ARIZONA DEPARTMENT OF TRANSPORTATION

REPORT NUMBER: FHWA-AZ87-244

TRANSPORTATION DATA MANAGEMENT

Phase I

Prepared by:
COMSIS Corporation
11501 Georgia Ave.
Wheaton, MD 20902

August 1987

Prepared for:
Arizona Department of Transportation
206 South 17th Avenue
Phoenix, Arizona 85007
in cooperation with
U.S. Department of Transportation
Federal Highway Administration

The contents of this report reflect the views of the authors who are responsible for the facts and the accuracy of the data presented herein. The contents do not necessarily reflect the official views or policies of the Arizona Department of Transportation or the Federal Highways Administration. This report does not constitute a standard, specification, or regulation. Trade or manufacturer's names which may appear herein are cited only because they are considered essential to the objectives of the report. The U. S. Government and the State of Arizona do not endorse products or manufacturers.

TECHNICAL REPORT DOCUMENTATION PAGE

1. REPORT NO.	2. GOVERNMENT ACCESSION NO.	3. RECIPIENT'S CATALOG NO.
FHWA/AZ 87/244		

4. TITLE AND SUBTITLE	5. REPORT DATE
Transportation Data Management	August 87
	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S)	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS COMSIS Corporation 11501 Georgia Avenue Wheaton, MD 20902	10. WORK UNIT NO.
	11. CONTRACT OR GRANT NO. HPR-PL-1-31-(244)
12. SPONSORING AGENCY NAME AND ADDRESS ARIZONA DEPARTMENT OF TRANSPORTATION 206 S. 17TH AVENUE PHOENIX, ARIZONA 85007	13. TYPE OF REPORT & PERIOD COVERED
	14. SPONSORING AGENCY CODE

15. SUPPLEMENTARY NOTES

Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration

16. ABSTRACT

Transportation agencies and other organizations need to differentiate between data and information. They need to start looking at information as a resource that must be managed and maintained.

In order for organizations to be effective, they must have the ability to design and manage integrated, functional, and diversified computer applications.

This study evaluates various information management systems which would be appropriate for medium size transportation planning organizations.

17. KEY WORDS	18. DISTRIBUTION STATEMENT
	Document is available to the U.S. public through the National Technical Information Service, Springfield, Virginia 22161

19. SECURITY CLASSIF. (of this report)	20. SECURITY CLASSIF. (of this page)	21. NO. OF PAGES	22. PRICE
Unclassified	Unclassified	104	

TABLE OF CONTENTS

Section		Page
1.0	INTRODUCTION	1-1
1.1	STUDY PURPOSE AND BACKGROUND	1-3
2.0	PHASE I	2-1
2.1	PRIORITY DATABASE APPLICATIONS	2-1
2.1.1	Information Management in Transportation Agencies	2-3
2.2	EVALUATION METHODOLOGY	2-5
2.2.1	Introduction	2-5
2.2.2	Minimum requirements	2-6
2.2.3	Identification and Ranking of Desired Features	2-7
2.2.4	Objective Features	2-7
2.2.5	Subjective Features	2-9
2.2.6	Rating System	2-13
2.3	APPLICATION OF THE EVALUATION PROCESS	2-16
2.3.1	Categories of DBS	2-21
2.3.2	Standardizing Analytical Abilities	2-23
2.3.3	Distinguishing Characteristics Among Databases	2-23
2.4	RECOMMENDATIONS	2-28
2.4.1	Database Package	2-30
2.4.2	Supplemental Software	2-32
2.4.3	Additional Hardware Requirements	2-32
2.4.4	Staffing Considerations	2-32
3.0	PHASE II - IMPLEMENTATION OF THE TRANSPORTATION IMPLEMENTATION SYSTEM	3-1
	Task 1 - Procurement of Materials	3-1
	Task 2 - Design of the Information System	3-2
	Task 3 - Development of System Modules	3-3
	Task 4 - Loading the Database	3-5
	Task 5 - Demonstrate System	3-5
	Task 6 - Provide Training	3-6
	Task 7 - Assess and Document Transferability	3-7
	Task 8 - Document System	3-7
	Task 9 - Final Report	3-8
	APPENDIX A - VENDOR TRACKING FORMS	
	APPENDIX B - QUESTIONNAIRE	
	APPENDIX C - SUMMARIES	
	APPENDIX D - GLOSSARY	

Prepared in cooperation with the U.S. Department of Transportation, Federal Highway Administration.

The authors wish to acknowledge Mark Peterson, Dave Wolfson, and Tom Buick of the Pima Association of Governments, Transportation Planning Division, and Robert Pike of the Arizona Department of Transportation for their contributions to and assistance with this project.

PREFACE FROM THE GEOGRAPHIC DATA BASE MANAGEMENT WORKSHOP

Alexandria, Va. March 2-4, 1983

"As local transportation agencies have increasingly acquired the technological trappings of computer automation, the problem of data management--or more properly, information management--is becoming a critical organizational issue. While we have acquired expensive computer hardware and software and have developed sophisticated models and analysis tools, the old adage of "garbage in, garbage out" has never been more true. As our computer equipment and analytical methods have increased in number and complexity, the quality and currency of local agency information has declined. The normal process of automation has been one of simply "computerizing" our manual records and forms with the all the duplication and inaccuracy inherent in these records, thus the process of automation can lead to an exponential proliferation of poor quality and redundant data."

"Until organizations begin to differentiate between data and information and start to look at information as an **organizational resource** that must be managed and maintained in support of functional operations, agency information will never be of any real value for planning and management, and information system managers will continue to play the role of data garbage collectors."

"The effective use of computers and information systems to support local transportation agencies is directly related to the ability of these organizations to design and manage integrated, functional, and organizational data bases that can be independently utilized to support multiple generic applications."

As part of the efforts of UMTA and FHWA to promote the effective use of new tools for transportation planning and system operation, this geographic data base management workshop has been organized to evaluate and synthesize recent experience with geographic data base management technology..."

1.0 INTRODUCTION

The above preface from an UMTA sponsored workshop on geographic data base management summarizes the need for improved information management among transportation planning agencies. It also provides us with a glimpse of the two-edged sword that this technology represents.

As the tools of technology become more affordable and attainable for Metropolitan Planning Organizations, there are a number of dangers. First, there is a danger that the tools which we seek to expand our efficiency and effectiveness will become a distraction and impediment to our true purposes (providing a continuing, comprehensive and cooperative process for management and development of the regional transportation resources). Second, the rising expectations of our public officials, of our citizens and of ourselves will surpass our ability to effectively incorporate these new tools and capabilities in the "3C" transportation planning process. Finally, there is the danger that we will be immobilized by the rapid advances in technology to the point that we fail to capitalize on the new found opportunities.

In addition to the dangers, there is also great potential. With the tools which are now within our reach, we will be able to more effectively communicate the meaning behind our analyses. There is the potential for rising beyond the routine of merely managing the tools of our trade towards using those tools to gain more insight into the relationships we seek to model and the development patterns we seek to control. There is the potential for providing a CONTINUING perspective of the region's direction

rather than disjointed snapshots; the potential to provide a more COMPREHENSIVE view of the implications of alternative actions rather than identifying impacts on parts of the system; and finally, there is the potential of providing a COOPERATIVE planning process, rather than separate organizations, each with "their" own data, working independently towards similar goals.

In authorizing the development of a Transportation Information System, PAGTPD and ADOT have sought to come to grips with the technological potentials and to integrate advanced information management tools into PAGTPD's standard practices. This choice was made with thought to the potential for greater efficiency and effectiveness in the conduct of PAGTPD's duties as well as the potential transfer of PAGTPD's experiences and tools to other MPO's in Arizona and throughout the country.

This report marks the conclusion of the Phase I effort and a major decision point for PAGTPD and ADOT. The questions concerning the feasibilty of the approach, the availability of the tools and the costs of the proposed system have been addressed; what is needed now is a confirmation of the committment and purpose which was evident at the start of Phase I.

The key question is whether the dangers of this newly accessible technology are more frightening or the potential more alluring.

1.1 STUDY PURPOSE AND BACKGROUND

The purpose of this study was the evaluation of the information management needs of the PAGTPD and identification and implementation of a strategy to allow PAGTPD to capitalize on advances in computer and information management technologies.

Initially, the project consisted of a sequence of tasks which included implementation of a Transportation Information System for PAGTPD. However, prior to the initiation of the project, it became apparent that too many uncertainties were involved and the intended scope of the database system inadequately specified to allow a practical plan for implementation. As a result, the project was redesigned to be accomplished in two phases. Specifically, Phase I was designed to:

1) identify those data items and activities typical of PAGTPD's operation which would most benefit from these technologies,

2) develop an evaluation methodology to identify the database package most consistent and compatible with the data management strategy, and

3) apply the evaluation methodology and produce a set of recommendations which would move forward that strategy and clarify the steps necessary for implementation of the information management strategy.

Phase II is to consist of acquision of the necessary hardware and software components, design and development of the application modules and the implementation of the Transportation Information System.

The identification of priority database applications and development of an evaluation methodology, Tasks 1 and 2 of Phase I, were previously documented in a joint memorandum. An updated

version of that memorandum is included as an attachment of this report and will be reviewed here only as necessary to provide context and identify modifications to the original documents.

2.0 PHASE I

2.1 PRIORITY DATABASE APPLICATIONS

Based on interviews and discussions with, and brief written statements from PAGTPD staff, eight applications of database functions and information management were identified. These areas, listed roughly in order of importance, are:

1) information management for the Travel Demand Forecasting System,

2) information management for socio-economic, project programming and non-model transportation data,

3) information processing for the transportation data,

4) statistical and other enhanced analytical capabilities applied to data in the database system,

5) graphical/mapping interface and display of the database,

7) maintenance of other miscellaneous sets of data, and

8) information exchange with outside agencies.

Primary among these applications is information management for the Travel Demand Forecasting System. Integral to the information management function are the information processing and reporting functions expected of the database system. Corresponding functions are also desired for the socio-economic, project programming and non-model transportation data maintained and utilized by PAGTPD.

The second most important function expected of the database system is the data library/data administration for data maintained internal to the database system, and for external datasets.

Aside from the transportation data management and processing functions, PAGTPD officials are most interested in applying advanced analytical capabilities to the data maintained within the database system. The main types of analyses tools are: spreadsheet capabilities, business graphics, statistics, and area and network mapping capabilities. In particular, the development of area and network mapping capabilities for both transportation and socio-economic data are highly desirable. A graphical interface to the database was also of interest, but such a capability is judged to be of lesser importance than other analytical capabilities. During the evaluation of database packages, the area mapping capabilities were also relegrated to a role of secondary importance. The potential for these capabilities was considered in the evaluation process.

Another area of interest to PAGTPD officials is the development of data exchange capabilities with outside agencies such as the Arizona Department of Transportation. While the most immediate potential for such data exchange is in regard to project programming information (data is currently exchanged via spreadsheet data files), routine data surveillance and information exchange in other areas could be developed. This area offers significant potential for advancing the coordination and cooperative aspects of PAGTPD operations. While PAGTPD officials recognized this potential, it was judged most appropriate to develop the in-=house capabilities of PAGTPD prior to pursuing this avenue.

PAGTPD staff also identified a limited number of other datasets which could be accommodated within the transportation

database system, e.g., accident information, transportation library. These datasets are typically relatively small in size, often having distinct processing requirements, and are also often independent of the primary transportation data elements to be maintained in the database system. For datasets fitting these characteristics, it is proposed that they not be integrated directly into the transportation database, but developed as separable databases serviced by a commercial database package interconnectable to the transportation database system.

2.1.1 Information Management in Transportation Agencies

In order to confirm the appropriateness of the information management strategy outlined in Task 1, several transportation agencies were contacted to review their current practices. Among the agencies contacted were Southeastern Michigan Council of Governments (SEMCOG), the Center for Urban Analysis (Santa Clara, CA), and the North Central Texas Council of Governments (NCTCOG)(Arlington, TX). Each of the agencies contacted has implemented transportation planning data management systems.

SEMCOG employs a hardware environment similar to that available at PAGTPD, but on a larger scale. Both TRANSPLAN and MicroVAX are utilized, and the current data management facilities depend on the use of SPSS and custom data transfer routines. SEMCOG has plans to develop an information management system based on a commercial database system, rather than continuing to rely on a less structured data management environment.

The Center for Urban Analysis employs a UNIX environment for their work and adopted a strategy which centers on the in-house

development of information management tools. This strategy has produced a graphical network editor, a geographic base file and a geographic information/mapping system, which are available either commercially or on a cost-reimbursement basis. The mission of CUA differs substantially from that of Metropolitan Planning Organizations (MPO), and adoption of their strategy is not practical for most MPO's.

The Center for Urban Analysis (CUA) is a unit within the Office of the County Executive (Santa Clara, CA) established to help local government agencies solve problems that have a geographic character and provide new problem-solving tools. The Center operates on a cost-recovery basis and its clients have included state and federal agencies, other local governments, and a variety of private and non-profit organizations. The Metropolitan Transportation Center and the Association of Bay Area Governments are the MPO's for the San Francisco region, including Santa Clara, and thus the CUA is not constrained by regular metropolitan planning typically carried out by MPO's.

The CUA is more of a research and development organization than a planning office. While it addresses some of the same problems as PAGPTD, its focus is determined more by its development objectives and the needs of its clients than the need to conduct the 3C planning process.

NCTCOG operates in an environment which also emphasizes in-house development of tools and procedures. A vastly customized adaptation of the Urban Transportation Planning System (UTPS) is used, as is TRANPLAN and MicroTrips. The staff is both large and

specialized with over 30 persons available to support the transportation forecasting system. Separate divisions handle the socio-economic data, project programming and monitoring functions. The forecasting section employs a series of programs developed in-house specifically to support the functions of that division.

One point which became clear in discussions with personnel at the above agencies is a recognition of the need for information management for the transportation planning process. While each of the organizations has pursued that goal in different ways, discussions with the staff verified that the use of a commercial database package as a base for development of PAGTPD's system would be the most efficient and cost-effective option. The use of a commercial database package would prevent a great deal of duplication of effort. Many of the different options and processes required by PAGTPD are included in most commercial packages and therefore would not have to be programmed by PAGTPD staff.

2.2 EVALUATION METHODOLOGY

2.2.1 Introduction

This section is designed to update and expand upon the evaluation methodology outlined in the Joint Memorandum for Tasks 1 and 2. The priorities and methodology of the evaluation process in this section were revised after reviewing the available literature, talking with software vendors, attending seminars and talking further with PAGTPD staff.

The section discusses the criteria which were used in reviewing and rating the different software packages. The format for the review process is basically the same as that developed in Task 2, encompassing minimum criteria, objective and subjective desirable features, and the numerical rating system.

2.2.2 Minimum Requirements

The basic requirement for all database systems (DBS) evaluated was the ability to operate on a MicroVax II under VMS in a multi-user, shared data environment. The DBS had to be relational, due to the nature of PAGTPD data and the flexibility afforded by the relational data model. The DBS had to have an interface with FORTRAN that would allow for development of customized programs utilizing data contained in the database system. This gives the programmer the ability to manipulate databases from within a FORTRAN program without having to reformat the data into FORTRAN readable form.

Another minimum requirement was that the DBS must be fully documented and actively supported by the vendor. Only those DBS' that come with complete detailed documentation including a user's guide, programmer's guide, and reference manual, and for which additional training is available, were considered in this evaluation. Most vendors charge an annual maintenance fee that includes technical support and software/documentation revisions. However, the level of service provided by these maintenance agreements varies from vendor to vendor. The minimum requirements for vendor support were telephone technical support and

problem solving, and access to all software and documentation updates.

2.2.3 Identification and Ranking of Desired Features

The previous section detailed the minimum requirements for the DBS. This section will discuss DBS features which are highly desirable to PAGTPD. They are divided into two categories, objective and subjective. Objective features are those that are easily measured and usually involve a yes/no answer as to whether the DBS has or doesn't have a feature. The subjective features focus upon the degrees of quality or completeness of the functions provided, and optional features of the DBS. For both the objective and subjective feature categories, it was judged appropriate to use a numeric scale (1 to 5) to allow for distinguishing among the levels of support for particular features.

2.2.4 Objective Features

The DBS chosen should, ideally, support some method of handling geographic data types. Methods include variable length text fields and bulk data types as well as built-in support for vector and matrix data. The file and record length limitations have to be large enough so as not to cause problems handling PAGTPD data. Data should not have to be tailored to fit the DBS package, the package should fit the data.

The DBS selected will have a "menuing" capability. Menu driven systems allow users with limited database experience to move easily within the database and perform complex functions with a few keystrokes. Many DBS' have a built-in menu generator

that allows the user to design menus on the screen. A system such as this would be the easiest to use in developing a complete information access system. Other DBS' have internal programming languages the user can employ to design menus, or require the use of a standard programming language such as FORTRAN. This would require a greater level of expertise than one would expect from the average user and would require greater resources to develop a menu-driven system.

Communications support provided by the DBS would allow the reception and transmission of data at all of the standard Baud rates supported by both the hardware and operating system. Also, the remote transmission device, such as a serial port, should be controllable from within a program written in the DBS programming language or in FORTRAN.

The DBS must provide basic mathematical (+,-,*,/) and statistical functions. The basic statistical functions should include means, minimum/maximum, standard deviation, and variance. Support of an expanded mathematical operator set, such as exponentials, natural logs, etc., within the DBS would be advantageous, due to their common use in transportation planning.

The DBS should provide a security system that is integrated into the active data dictionary. This system should allow the system manager to allow or deny access at the database, table and field levels.

An on-line help system will be an integral part of the DBS selected. This allows users to receive information concerning command syntax, procedures, etc., without having to refer to the

written documentation. The on-line help function should not, however, replace the written documentation as the final and most complete resource. The help function should have the ability to updated and modified by the user to include help for user-defined procedures that are added to the DBS.

The first year cost and the five year cost of the DBS were both considered because of annual maintenance and licensing fees charged by vendors. A DBS with a relatively inexpensive purchase price may have an exhorbitant annual fee that makes its cost prohibitive over an extended period of time.

2.2.5 Subjective Features

The subjective features have been divided into two groups, based upon their importance in relation to the overall goals of PAGTPD. These groupings were reflected in the numerical rating system, with the features in Group One receiving higher weights than the features in Group Two.

Group One

The FORTRAN interface supplied with the DBS may have different levels of completeness. The interface may only support basic access to data within the database through its FORTRAN interface, or the interface may provide high level functions comparable to those directly executable from within the database system.

The data and file manipulation abilities of the DBS can be represented in many different ways. Does the DBS allow subsets of virtual tables to be joined with other tables and subsets? Are variable length fields available? Questions such as these

and others involving the use of the database must be answered by the DBS.

The DBS must have the ability to store repeated queries so that they can be recalled without having to be re-entered.

The database query language must be complete enough to allow the use of computed values and multiple criteria in select commands and allow queries and subsetting of virtual tables.

The DBS must have a recovery system thorough enough to provide full backup and recovery in the event of corruption or destruction of the database.

The DBS must provide a report writer that permits the production of customized forms and formatted output. In the best case, the report writer would include a "screen painting" system that lets the user design a form without intensive programming. The report writer must also have mathematical and summary operators to calculate totals, subtotals, means, etc., at break points throughout a report.

Simple business graphics should be easily generated from the databases. The graphics (pie charts, bar graphs, line graphs, scatter plots, etc.) specifications should be retrievable and should be applicable to virtual tables.

The DBS should support a full spreadsheet function that has the ability to draw data from the database. The size of the spreadsheet must be sufficient to allow a substantial sized table to be loaded.

Advanced statistical functions such as cross tabulations, regression, correlation and analysis of variance should be

supported.

The DBS should support geographic based data with features such as table look-up and aggregation. Table look-up would be used to determine relations between geographic based data elements (e.g., which nodes are in a certain zone). An aggregation function would allow for summing of data over a defined area (e.g., summing the VMT for all links in a zone would produce the VMT for the whole zone). It ideally would support a graphical mapping user interface to identify areas of interest.

The speed of storage/retrieval of the DBS is most easily evaluated by the use of benchmark testing. Efforts to obtain an objective assessment of the relative speed of different database packages were frustrated by incomplete, poorly documented or incompatible benchmark assessments available from the vendors. Some DBS packages have a "fast load" capability that loads data quickly, but does not perform error or integrity checking.

The DBS data capacity should be sufficient enough that there is never chance of PAGTPD's data exceeding that capacity. The data should not "outgrow" the system after a few years. This would require the purchasing of another, more capable, larger DBS in a short amount of time.

The method by which the DBS adds/deletes attributes is a time critical factor. If the database structure can be modified when the database is live, there is no time loss. If, as with many DBS', the data must be off-loaded before the structure is changed, the time cost could be significant and discourage modification of the database to best suit the information processing requirements.

Group Two

The DBS should support importation of different data formats (i.e., ASCII, EBCDIC, DIF, dBASE III, Lotus 1-2-3). PAGTPD receives data from a variety of sources and applications, and if conversions must be performed each time new data is input, the productivity and effectiveness of the system is diminished.

On-help facilities built into the DBS should be thorough enough to answer basic questions about the purpose and syntax of commands, so the user is not continually referring to the written manuals. The on-line help function cannot and should not replace the written documentation as the final and most complete reference. The on-line help function should have to flexibility to allow the user to add his own explanations of custom procedures.

The DBS must have an operating system interface that lets users call other programs or applications and use operating systems functions from within the DBS.

The DBS should support user and query logs that would tell the system manager which users are on the system the most, and which databases and procedures they are using. The logs would also be an aid towards improving the system's efficiency security by informing the system manager of tables, fields and procedures most frequently utilized.

The amount of standardization of the DBS query language is an indication of the ease of portability to other systems. This would decrease the amount of specialized training needed and would allow less experienced users to access the system. An

additional benefit is in terms of the transferability of the system to other MPO's.

2.2.6 Rating System

Before any of the DBS' were rated quantitatively, they were all rated qualitatively (Figure 1). This guaranteed that each of the DBS under consideration met the minimum requirements specified by PAGTPD and were among the higher quality database packages before being numerically rating, therefore saving time and effort. The numerical rating system used was not meant to be a formal, precise or complete method for differentiating between DBS packages. It is meant to be used in conjunction with the verbal evaluations to provide a complete review of the available packages. The evaluation process makes use of both the quantitative and qualitative assessments of the database packages. The numerical scores of the database packages are used to focus the evaluation and to identify various groupings of packages. The most important aspect of the evaluation process is the qualitative and necessarily subjective assessments. A numerical scale from 0 to 5 was used, with the following assigned values:

5 - Excellent, feature exceeds requirements;
4 - Good, feature meets requirements;
3 - Adequate, acceptable;
2 - Poor, limited capabilities;
1 - Inadequate;
0 - Feature is not present.

Because some features were more deemed more important than others by PAGTPD, a weighted rating approach was used. The weights range from two to four, and indicate the importance of the

Figure 1
NUMERICAL RATING SYSTEM

	Rating	Weighted Rating
Objective Features		
Support Geographic Data Types	___ x 4 =	___
File and Record Specifications	___ x 3 =	___
"MENUING" Capability	___ x 3 =	___
Communications Support	___ x 3 =	___
Basic Statistical and Mathematical Functions	___ x 3 =	___
Security System	___ x 2 =	___
On-Line Help System	___ x 3 =	___
First Year Cost	___ x 2 =	___
Five Year Cost	___ x 2 =	___
Subjective Features		
Group One		
Completeness of Programming Language	___ x 3 =	___
File Manipulation Ability	___ x 3 =	___
Store Repeated Queries	___ x 3 =	___
Completeness of Query Language	___ x 4 =	___
System Recovery	___ x 3 =	___
Formatted Output	___ x 4 =	___
Graphical Interface	___ x 3 =	___
Business Graphics	___ x 3 =	___
Spreadsheet Function	___ x 3 =	___
Advanced Statistical Functions	___ x 3 =	___
Geographic Based Data Functions	___ x 4 =	___
Speed of Storage/Retrieval	___ x 4 =	___

	Rating	Weighted Rating
Capacity for Data	___ x 3 =	___
Ease of Adding/Deleting Attributes	___ x 4 =	___
Group Two		
Support Different Data Input Formats	___ x 2 =	___
Completeness of On-Line Help Facility	___ x 2 =	___
Operating System Interface	___ x 2 =	___
Query Log	___ x 2 =	___
User Log	___ x 2 =	___
Portability to Other Systems	___ x 2 =	___
	Total (Max=420)	___

particular feature in relation to PAGTPD's goals. The highest possible score was 420.

2.3 APPLICATION OF THE EVALUATION PROCESS

The first step in the evaluation process was to identify DBS available for the MicroVax II that met the minimum criteria. PAGTPD supplied the consulting team with a review of DBS available for DEC equipment found in a 1986 issue of Digital Review magazine. The review listed the major attributes of the many different DBS available for VAX equipment. The criteria used to select packages from those listed in the review were that the DBS could run them in a VMS environment, were relational database systems, and could support FORTRAN through a programming interface. Those vendors whose packages met those criteria were sent letters informing them of the consulting team's mission and requesting any and all information regarding their DBS packages.

The content of the vendors' responses varied a great deal. Some sent brochures consisting of a few pages of narratives, while others responded with highly detailed specifications and examples. The literature was reviewed and where necessary, vendors contacted for additional information. The field of packages was narrowed, based mainly on the criterion that the DBS be fully relational and offer the facilities necessary for development of a menu-driven, extendable information system.

Many packages that claimed to be relational turned out not to be, when following the strict definition of the concept. The remaining vendors were then contacted by phone to establish a

relationship so future inquiries would be easier. The purpose of the project was explained fully and more detailed information was requested.

During this process a tracking system was established by the consulting team to monitor the responses from and contacts with the different vendors. This system let the consulting team easily know the status of each package being considered. A checklist that indicated if a desired feature was present in the DBS was also included in the tracking system. If a DBS was eliminated from contention, the tracking system contained the reason for elimination and the date the decision was made. The tracking system also contained the vendor contact person's name and telephone number for quick reference. An example of an entry for this tracking system is given as Figure 2. Entries for the remaining finalists among the database packages are listed in Appendix A.

After the latest requests for information were received and reviewed, it was decided by the consulting team that a more standardized format for comparison of packages was needed. This led to the development of a questionnaire (Appendix B) that was used to guarantee that each vendor was asked if his package met PAGTPD requirements and contained the features discussed previously in this report.

The original list of database packages considered included approximately 44 packages from various vendors. This list of database packages for the MicroVAX was compiled from an article in Digital Review and from database packages suggested by PAGTPD.

Figure 2

ARIZONA DEPARTMENT OF TRANSPORTATION
PIMA ASSOCIATION OF GOVERNMENTS TRANSPORTATION PLANNING DIVISION
TRANSPORTATION DATABASE PROJECT

Database Name: Accent R Phone: (408) 257-7700

Vendor: National Information Systems, Inc.
Representative: John Enyedy

Address: 20370 Town Center Lane, Suite 130
City: Cupertino State: CA Zip: 95014

First Year Cost: $20,140.00
Annual Cost after First Year: $2,600.00

Remark 1: Vendor claims a 2-1 speed advantage over Oracle & Ingres
Remark 2: Interfaces to spreadsheet, statistics, graphics & word processing
Remark 3: HLI a recent development for the VAX implementation
Remark 4: Price of graphics/spreadsheet from different vendor included in cost above
Remark 5: !! REQUIRES database be off-loaded to add/delete attributes!!

A review of these packages produced a total of sixteen packages which appeared to meet the minimum criteria established for the transportation database system and which warranted further further consideration. Additional information about these sixteen packages was requested from the appropriate venders through both written and phone requests. A comparative assessment of these packages then eliminated ten of the packages from further consideration. Packages eliminated included those which upon closer inspection were not suited to the intended application, did not meet all of the minimum criteria or were substantially less capable than the other packages under consideration. The six packages which remained were scrutinized in detail and assessments of their features are summarized later in this report.

At this point, the field had been narrowed to six packages: Boeing RIM, Ingres, Empress/32, Oracle, Accent R, and InFoCen. The vendor representatives of these six packages were contacted to answer the questionnaire. Most responded over the telephone, but one or two requested that the questionnaire be mailed to them. The results of the questionnaire were added to the tracking system.

At the same time as the literature review and information compilation was taking place, other information sources were explored. Members of the consulting team attended seminars sponsored by the vendors. This allowed the team to see the DBS packages "live" and question the vendors in person.

A variety of users were contacted to get their impressions of the two final packages, Ingres and Oracle. Some of the users

had evaluated both products prior to making a decision, while others had only casual exposure to the competing products. The users of each package thought that the one they had chosen was the better of the two, but not always for the same reasons. Some of the varying opinions are presented below:

- Ingres was favored because of its ease of use for novices and flexibility. Oracle also received high marks for its user friendly environment.

- In the area of customer support, Ingres appeared to have faster and more accurate responses to user problems. One user indicated that if the Ingres representative did not know the cause of a problem, he would help devise a way to work around the problem until it could be corrected.

- Oracle was rated higher for its documentation, and both packages were praised for their training.

- In the area of performance, the users of each system claimed superior speed. However, one user stated that the differences in performance were negligible, since as each new revision of Oracle came out, it would slightly outperform the current version of Ingres, and vice versa. In addition, the performance and speed of both packages are limited by the VAX architecture.

- The report writer feature of Ingres was rated higher than Oracle's by people who had used both. Oracle's report writer was reported to be awkward to use and not particularly user friendly for beginners.

- The Ingres graphics option was said to have caused severe performance problems for all users on the system even if in use by only one person. One user, however, thought that in a multi-user environment, Ingres was the better performer.

- For updating and changing database structures, Ingres was rated highly. For the initial design and implementation of a database system, Oracle received good reviews.

Overall, the users perceived the package they used as the best. Even those people who had used both packages at one time or another considered the one they are currently using as the

better one. Most of the persons contacted were long time users of database systems and were in charge of the design, implementation, and maintenance of the DBS for their company or agency. In summary, the review of user experience with the two packages tended to reinforce the comparability of Ingres and Oracle in terms of features and performance. Each was viewed as better by those using the product and no substantial differences or deficiencies were identified.

Another source of information that was explored was the independent reviews and tests of the different DBS'. Digital Review reviewed one of the packages quite thoroughly and provided useful benchmarks concerning processing and storage speed. Finding comparable reviews and benchmarks for the other packages created a problem simply because there were none available. DECUS, a DEC user's group was also contacted for their opinion on the final packages.

Summaries of the final six packages were developed, along with numerical ratings, and were forwarded to PAGTPD for their review and comment. The narrative summaries of the final six packages are given in Appendix C.

2.3.1 Categories of DBS

During the review process, the consulting team discovered that there were some inherent differences between packages that made them either better or less suited for certain types of jobs. The differences usually centered around the purposes of the DBS and the type of data each was designed to handle. Some packages were designed specifically for scientific uses and contained

minimal advanced menuing capability and other user-friendly attributes. Other packages emphasized their text handling abilities. These systems are well suited for library and research database development, where searching and locating based upon a string of text is critical. Many packages are more general in purpose and may be categoried as business oriented databases. This category varies considerably in the features and conveniences offered. Some packages maximize data access by using system files but suffer in terms of the data manipulation capabilities or efficiency of operation. Most packages, however, use a propRietary file format to house the database.

The main difference between DBS' that became a factor in selecting a package for PAGTPD was query language. Most vendors base their query language on the proposed ANSI's Standard Query Language (SQL), which is theoretically the industry standard. However, it is not a standard as on the level of FORTRAN, for example. Most vendors, when employing SQL as their query language, usually extend it and add more complex and powerful commands. A number of vendors, however, choose to develop their own proprietary "Non-Procedural" Query Languages. Therefore, a major decision for the study team was choosing between SQL and non-SQL based packages.

The major strong point in favor of SQL based DBS is that much of the knowledge can be transferred to and from other systems. This could save time and training when PAGTPD installs the system. During the evaluation process, both types were reviewed and were represented in the final six DBS'.

2.3.2 Standardizing Analytical Abilities

In order to make comparisons more fair and consistent, the packages had to be standardized so that all contained the same basic analytical abilities. The most important difference between DBS' in the final six was that one contained a spreadsheet function and business graphics, another contained only graphics, and four packages contained neither spreadsheet nor graphics. It was decided to include a spreadsheet package from a separate vendor which included graphics and spreadsheet functions comparable to that available in the most fully featured package in the price of the other four packages containing neither spreadsheet nor graphics.

2.3.3 Distinguishing Characteristics Among Databases

Query/Data Manipulation Language

Of the packages which were evaluated in depth, several features were useful in distinguishing amount the packages. The first distinguishing feature is the query/data manipulation language provided to manipulate the database. Three of the packages offered complete SQL implementations while the remaining packages used proprietary languages. The proprietary languages used by the other packages were judged to be comparable in terms of their data handling capabilities, this aspect of the SQL implementations did not provide any discernable advantage.

Portability

The availabiiity of SQL is worth noting for two reasons. First, given the desire to develop a prototype system for the

Transportation Database, a system which is developed using a standard language is inherently more portable to other environments where the database package and/or the operating system may be different. Additionally, the development or acquisition of staff skills is facilitated by the use of a widely adopted query/data manipulation language. Thus, the packages offering SQL were judged to have an advantage in the PORTABILITY evaluation criteria (the availability of the packages for other operating systems was also considered in assigning a score for PORTABILITY).

FORTRAN Language

A second advantage of the packages offering SQL implementations is that each of them also provided superior FORTRAN INTERFACES for the development of custom programs using the database. All of the packages considered in the final round of the evaluation offered a Host Language Interface (HLI) which provided basic access to the database. Typically, the procedures provided allowed a table to be opened and closed, data records to be inserted, updated, deleted and transmitted to the program. Other basic features such as access to the data dictionary and indicies may also be provided. The packages offering SQL went one significant step beyond this point by providing the option to embed SQL statements in the source program and pre-compiling the SQL statements in to procedure calls. This option significantly enhances the ability to develop sophisticated applications. One SQL statement may be the equivilent of several hundred FORTRAN statements and allows the programmer to concentrate his attention

on the user interface and applications rather than the details of data manipulation. This was considered a significant advantage and was reflected in the FORTRAN INTERFACE scoring.

Virtual Table Support

Another notable benefit of the SQL implementations is the full support for views of the database. One promise of relational database systems is a simple consistant view of the data presented in the manner which is most useful to the user while reducing data redundancy and manipulation requirements. Support for views or virtual data tables is one of the mechanisms used to provide those benefits.

A view can be generally described as a table comprised of fields (attributes) from one or more tables of the database which does not physically exist but which the database is able to present to the user as a table for manipulation or query operation. In this sense it can be considered a virtual table. The database packages considered offered various degrees of support for views of the database. All offered some degree of support, however, this support was not always comparable to true tables in the database. In many cases the database allowed query operations or joins using the view but required the creation of a physical table for reporting, updating or exporting. Such a requirement negates some of the benefits of virtual tables and places additional strains on data integrity, data storage and manipulations requirements.

Restructing the Database

In considering the potential need to restructure the database by the addition/deletion of fields of a table, another distinguishing feature was identified. Most of the systems considered required off-loading of data or other reduced functioning of the database to restruct the database schema. While such a process should not be frequently necessary, having to shut down the database, off-load the data, revise the schema and reload the data would be a significant inconvenience. The SQL products consistently scored better in this category than did the non-SQL products. The SQL databases were all able to update the structure of a table with the database fully active. The non-SQL databases offered no directly comparable capability.

On closer examination, it appears that those packages not offering this explicit capability could simulate this capability through a series of commands including making a copy of the table of interest. This was considered a less efficient than direct support for dynamic restructing of the database and this assessment was reflected in the scoring for the corresponding evaluation criteria.

Ease of Use

This evaluation criteria is comprised of several considerations. The ease of use is primarily concerned with the interface provided to occasional users of the database although the facilities provided for regular users was also considered. The availability of a menu driven environment was an explicit desired of the PAGTPD. Ideally, this environment would provide

complete access to functions within the database system as well as to custom programs and operations. The facilites provided in this area ranged from spartan to extensive. The more extensive packages included an applications generator (a tool for building custom programs or reports) as well as menu-driven query, report writer and data manipulation facilities.

Enhanced Analytical Capabilities

In addition to the management of transportation planning data maintained by PAGTPD, the Transportation Database should serve as a central focus of analytical and reporting activities. The inclusion of enhanced analytical capabilities within the database environment was an explicit desire of the PAGTPD staff and officials. The specific areas of interest included: business graphics, spreadsheet functions, advanced statistics and geographic information system functions. Only in the case of business graphics (pie, bar, line and text graphs) and spreadsheet functions were these services offered as an integral part of the database package. Oracle was the only package considered which provided a built-in spreadsheet. A few packages in addition to Oracle provided built in business graphics. In all cases, the database packages reviewed provided a means of sharing data with external programs but generally did not provide the same level of functionality or ease of use. Where the business graphics or spreadsheet function was not provided within the database system, a spreadsheet/graphics package from a seperate vendor which provided comparable capabilities was used to supplement the database package. Because a less integrated environment

was provided, those packages which provided the desired functions within the database system received a higher score in this category. In the case of one or two packages, minimal graphics capability was provided but the functionality was substantially less than that available from other systems. Those packages were treated as though the function was not available within the system.

All of the packages would provide substantially the same access to an external statistical package. The data would have to be extracted from the database and written to an external file in a format acceptable to the statistical package. A very slight advantage was seen for SQL based systems in that the two leading statistical packages, SPSS and SAS, provide a interface to another SQL based product (IBM's DB2). The potential exists for this interface to be extended by the packages' vendors to support other SQL based database packages. This potential was not judged sufficient to warrant a difference in the scoring of the packages being considered.

2.4 RECOMMENDATIONS

The results of the rating of the database packages are summarized in Table 1.

Based on the evaluation conducted in Phase I, there is conclusive evidence that the goals for the PAGTPD Transportation Database System can be effectively achieved using Ingres as the basic framework. Ingres fulfills the basic requirements specified for a database package. The complete system also affords a number of features which are desirable in the creation of the

Table 1

ASSESSMENTS OF DATABASE PACKAGES FOR PAGTPD

Evaluation Criteria	Weight	Accent R	Boeing RIM	Express/32	InfoCen	Ingres	Oracle
Support Geographic Data Types	4	0	1	2	2	0	0
File & Record Specifications	3	5	3	5	5	3	3
"Menuing" Capability	3	3	1	3	3	5	4
Communications Support	3	3	3	3	3	3	3
Basic Statistical & Mathematical Functions	3	3	4	3	3	3	3
Security System	2	0	3	4	3	3	4
On-Line HELP System	3	3	3	3	3	3	3
First Year Cost	2	3	4	4	3	3	2
Five Year Cost	2	3	4	4	3	3	2
Subjective Features-Group One							
Completeness of Programming Language	3	3	3	4	3	4	4
File Manipulation Ability	3	3	3	3	3	3	3
Store Repeated Queries	3	3	3	3	3	3	3
Completeness of Query Language	4	3	3	3	3	4	4
System Recovery	3	0	3	3	3	3	3
Formatted Output	4	3	3	3	3	3	3
Graphical Interface	3	0	0	1	1	3	2
Business Graphics	3	2	2	2	3	3	3
Spreadsheet Function	3	2	2	2	2	2	4
Advanced Statistical Functions	3	2	2	2	2	2	2
Geographic-based Data Functions	4	0	0	1	0	3	2
Speed of Storage/Retrieval	4	3	3	3	3	3	3
Capacity for Data	3	4	3	4	4	3	3
Ease of Adding/Deleting Attributes	4	0	2	3	0	3	3
Subjective Features-Group Two							
Support Different Data Input Formats	2	3	3	3	3	3	3
Completeness of On-line Help Facilities	2	3	3	3	3	3	3
Operating Systems Interface	2	3	3	3	3	3	3
Query Log	2	3	0	3	3	3	3
User Log	2	3	0	3	3	3	3
Portability to other systems	2	0	1	2	1	3	3
Summary Evaluation		186	195	241	217	247	241

Transportation Database System. Notable among the facilities provided are: use of a standard/portable query language; menu-driven access to the data manipulation, query, report, and business graphics components; a sophisticated interface to the FORTRAN programming language; and the documented ability to extend the environment to include tight integration with a geographical information system (including a graphical interface) and the flexibility and extendability to allow growth of the system to meet PADTPD's future needs.

2.4.1 Database Package

Ingres was selected as the database package which was best suited as the basic "engine" for the Transportation Information System. Among the characteristics which led to this conclusion were the level of sophistication of the data manipulation and query functions provided, the ease of use for occasional users, the facilities available for program development and the potential for an integrated reporting and analysis system.

The data manipulation facilities lie at the heart of the benefits of database systems and Ingres is unsurpassed in the capabilities provided. The query facilities are similarly advanced and supplemented by an interface which makes much of the database's power accessible to occasional users. The facilities provided for program development are matched only by that of Oracle and provide significant advantages for the development of an easy to use yet sophisticated information management system. Ingres also offers the potential to develop a thoroughly integrated geographic information system at a later date building

upon the basic system developed in Phase II of this project. The only comparable system lacks the same level of demonstrated potential and would cost approximately twenty thousand dollars more over a five-year period. This selection was reinforced by the assessments given by Ingres users contacted (see section 2.3).

2.4.2 Supplemental Software

In order to provide the level of analytical functions desired, it is necessary to supplement the acquisition of Ingres with the purchase of a spreadsheet package and a statistical package. The spreadsheet is intended to provide a platform for less structured analysis and exploration of the data while the statistical analysis is required to provide for model development and assessment as well as facilities requiring more sophisticated data analysis. These packages will be coupled to the database to provide a convenient yet comprehensive platform for much of the analysis conducted by PAGTPD. Although evaluation of spreadsheet and statistical packages was not in the scope of the Phase I effort, initial investigations suggest that appropriate packages can be readily identified. For the statistical component, either the Statistical Analysis System (SAS) or the Statistical Package for the Social Sciences (SPSS) would be well suited to complement the database package.

2.4.3 Additional Hardware Requirements

The current hardware is sufficient to support a limited Transportation Information System. The primary limitation of the existing hardware environment is in terms of fixed disk storage.

Based on current usage of the MicroVAX, the mass storage facilities should be expanded to allow for full development of the system. A minimum of 60 megabytes of fixed disk storage should be made available for the transportation database system.

2.4.4 Staffing Considerations

The development and operation of a Transportation Database System has implications beyond the acquisition of additional hardware and software. Specifically, implementation of the Transportation Database System will impact staffing of PAGTPD. The first impact will likely be an expansion of staff responsibilities. The administration of the transportation database and maintenance of the system will require time and attention beyond that which can be afforded at the current staffing level. This increase in staff responsibilities will be offset to some degree by the increased efficiency expected to result from implementation of the transportation database system. While it is expected that the increased efficiency of the PAGTPD staff will more than offset the staffing requirements for the database system, it is important that neither the maintenance of the transportation database nor the fulfillment of PAGTPD planning responsibilities be neglected.

For this reason, it is recommended that either a half-time or a full-time staff position be created. The administration of the Transportation Database System should be made a half-time responsibility staffed from PAGTPD's current full-time personnel. The newly created position would be used to fill the void left by reassignment of personnel. After the transportation database has

been operational for a period of twelve months, a re-assessment of staffing levels at PAGTPD should be conducted and adjustments made as necessary.

The degree of data integration, the enhancement of analytical abilities, the expansion of the staff's ability to respond to information requests and to move effectively to present the results of their analysis afforded by the proposed system would serve as a model for other MPO's in Arizona and across the county. Further, the system developed will extend the capabilities of PAGTPD staff in meeting the organization's mandate. With the tools provided, PAGTPD staff will be better able to convey the meaning behind the numbers as well as to more efficiently meet PAGTPD's responsibilities.

3.0 PHASE II - IMPLEMENTATION OF THE TRANSPORTATION IMPLEMENTATION SYSTEM

One objective of the Phase 1 effort was to further develop the scope and orientation of the Transportation Information System to be implemented in Phase II. The refined scope of the Transportation Information System is presented in the documentation for Task 1 of Phase I. That information is not repeated here but is incorporated by reference. This portion of the Phase I Final Report outlines the activities contemplated under Phase II, the implementation of the Transportation Information System.

This Phase will encompass several tasks, ranging from the acquisition of materials to the final delivery of the system and documentation. The implementation of the information system will emphasize modular system developement techniques both to aid in the development process and to provide for future expansion of system capabilities. The specific activities in the development of the system are discussed below.

Task 1 - Procurement of Materials

The first activity in Phase II will be the procurement and installation of the selected database package and its options. Concurrent with the acquisition of the database package will be the acquisition of a spreadsheet and statistical package to complement the analytical capabilities of the database package. The specific database options to be included in the procurement will be: FORTRAN Host Language Interface, Ingres/Menu, VIFRED (Visual Forms Editor), Report-By-Forms, Query-By-Forms, Graph-By-Forms (or VIGraph), Report Writer. Final selection of the spreadsheet and statistical package will be developed through

consultation with PAGTPD and the database vendor. If not previously acquired, additional hard disk capacity for the MicroVAX will also be obtained at this time.

Task 2 - Design of the Information System

Acquisition and installation of the basic materials will be conducted concurrently with the development of the database schema and formal specification of the standard operations and reports to be produced. This specification will form the basis for development of the program modules later in this task. Development of the database schema will include a comprehensive listing and formal definition of the data items to be integrated into the transportation database, identification of data items which are to be cataloged within the database but maintained separately. This information will be incorporated in the documentation of the database and also serve as a consistency check during implementation of the database.

The specification of standard operations and reports to result from implementation of the database system will complete the design of the information system. Among the items to be included in the operations' and reports' specifications are:

- o full integration of NEDS for network editing and display;
- o tightly coupled data transfer routines between the database, TRANPLAN, the spreadsheet and statistical packages;
- o design of standard reports for trip generation, trip distribution, mode split, and trip assignment;
- o design of standard reports for comparing alternative assignments, land-use scenarios, etc. at each stage of the travel forecasting process;

- o specification of modules for standard application of TRANPLAN programs including extraction of selected data from the database, invocation of TRANPLAN programs, transfer of data from TRANPLAN to the database and application of appropriate reporting modules;
- o Design of modules for managing socio-economic data, project programming data, air quality monitoring information and traffic counts as appropriate; and
- o Design of modules for reorganization of data for different zone/district reporting systems.

The reports referenced above will include numeric, statistical and graphic summaries. The specification of the components will include identification of data flows, sketches of textual/numeric summaries, and selection of statistical summaries and graphic formats to be applied. It is these components which will comprise programmed options to be incorporated in the Ingres menu system.

Task 3 - Development of System Modules

The development of the modules comprising the Transportation Information System will fully utilize modular programming techniques and will employ systematic measures to ensure the attainment of design goals, programming, documentation and quality assurance standards. As a first step in the application of the these documentation standards, format and programming guidelines will be established prior to any program development work. The user interface will be designed to emphasize clarity and consistency with the Ingres user-interface. Logical grouping of modules will be developed and the basic system of menu options specified. As each screen or menu is developed, one or more HELP screens will be developed to provide on-line assistance to users

in selecting among the options and clarifing any information requests.

Modules will be developed using the Ingres program development tools where appropriate, and a combination of FORTRAN and embedded SQL elsewhere. The use of SQL rather than QUEL will serve to maximize transferability as well as emphasizing the use of accepted standards. Components of the information system will be developed as self-contained modules with well-documented interfacing with other modules and the main program. This design philosophy will aid in the development of modules as well as in the testing, integration and review of components.

Rather than deliver a massive, completed system to PAGTPD for acceptance, we propose to first deliver the menu system (the shell of the information system) for initial acceptance and to deliver each module separately for review and tentative acceptance by PAGTPD. As each module is completed, it will be integrated with the menu system and delivered to PAGTPD complete with documentation and extensions to the HELP system. This will facilitate review and testing by PAGTPD as well as offering early feedback to the development of other modules. Final acceptance of the information system by PAGTPD will not occur in this task, but will follow integration and installation of all modules, demonstration of the system, training of users and delivery of final documention.

Task 4 - Loading the Database

In order to facilitate the modular development of the information system, loading for the database will proceed in two steps. First, following installation of the database package, standard data items will be loaded onto the database both to exercise the database schema as well as to provide a well defined set of real world data for testing of program modules. This initial loading of the database system will not be used by PAGTPD for its regular activities but will be utilized for testing and application development purposes.

The second loading of the database will follow initial acceptance of the final program module by PAGTPD. This loading will utilize the lastest version of PAGTPD planning and forecast data and will provide a production level data for demonstration of the system and training of users prior to final acceptance by PAGTPD. This will also ensure the delivery of a complete tested and operational information system at the completion of Phase II.

Task 5 - Demonstrate System

The demonstration of the system is intended to accomplish several goals. First, the demonstration should serve to exercise the capabilities of the system and visibly document the achievement of design goals. Second, the demonstration may serve to reveal any rough edges in the system and user interface. To accomplish both goals, we propose to conduct a two-tiered demonstration of the system. An initial demonstration will be made to the project steering committee while the system is in the "draft" stage. Based on comments received, the system will be modified as

necessary to address concerns raised. The "final" system will then be demonstrated again as part of the training package provided in Task 6. As part of this demonstration, a third goal will be achieved. The training may serve to identify concepts or applications which are not clearly communicated in the training sessions and draft user documentation. Feedback from this training session would then be used in finalizing the user documentation giving special emphasis to areas which were unclear in the training session and draft user documentation.

Task 6 - Provide Training

Training in the use of the system will take place at two levels. A general training session will be conducted by the study team to acquaint "casual" (regular) users of the system. It is also proposed that an individual at PAGTPD be selected for training as a "expert" user. This individual will participate in up to a week of training in the use of the database package offered by the software vendor as well as receiving training specific to the information system. After the conclusion of the contract, the expert user would be responsible for assisting others in the use of the system, answering user questions and possibly transferring the system to other MPO's. The expert user would participate in the general training session both to identify him/her as an expert in the use of the system as well as to provide additional exposure to the system prior to routine use. Similarly, the individual who will be responsible for system maintenance would be closely involved in initial design and in the documentation of the system.

Task 7 - Assess and Document Transferability

One of the goals of this project is to produce a system that can be utilized at other MPO's (MAGTPD and YUMA MPO). This transferability will be accomplished by judiciously selecting a data base system that does not limit the operating environment. Modular development will also allow changes to be implemented to accommodate special needs at each agency. Documentation of the portions of the system that would need modification would be accomplished as part of this task.

Task 8 - Document System

The documentation of the system will be an ongoing task in Phase II. As part of Tasks 2 and 3, the functional specifications will be fully integrated and the user interface defined. Modules will be developed using a top-down, structured approach with tightly defined interfaces between modules. Each module will be documented in functional terms (identifying the module's effects on the system) and operational terms (the specifics of how the module operates). Interdependencies among modules will be clearly specified in both the external programmer documentation and the source code. The source code will be liberally commented to assist in initial debugging and later maintenance of the system.

A standard documentation outline will be prepared to aid in consistency and completeness in documenting program modules. This outline will contain a text description of the module detailing the purpose, basic operation, files used, and procedure dependencies. The module's details will then be documented by

providing a pseudo-code description of the module's operation, reproductions of data input and output screens, and details of the file format used. Further global and local variables used in the module will be identified as will be temporary and permanent files accessed by the module. All procedures called by the module will be identified and the commented source code reproduced. Any unusual or "tricky" data manipulation will be identified and further explained in an appendix to the module documentation. It is expected that modules will be limited to approximately 10 double-spaced pages of source code. Operations which cannot be accomplished within this limit will be broken into modules meeting this limit to aid comprehension and potential modification at a later date.

Expansion capabilities will be designed into the system in three primary ways. First, the modular design of the system will serve to "insulate" the system from unintentional side effects. Second, a system of fundamental routines accessible to all modules will be developed to ease development and to standardize the system's operation. Third, the process necessary to expand the system by adding new capabilities will be fully explained in the programmer documentation.

Task 9 - Final Report

This report will incorporate the various elements of the project in a single document. This will include (as a minimum):

- research design,
- all phases of the data system development,
- software description,
- recommendations for future enhancements, and
- complete source code fully commented and annotated.

A draft final will be produced (12 copies) for ADOT review. A final report will incorporate review comments and a camera-ready report submitted.

APPENDIX A

Arizona Department of Transportation
PIMA Association of Governments Transportation Planning Division
Transportation Data Base Project

DataBase Name: Accent R PHONE: (408) 257-7700

Vendor: National Information Systems, Inc.
Representative: John Enyedy

ADDRESS: 20370 Town Ctr. Lane, Suite 130
City: Cupertino State: CA. ZIP: 95014

First Year Cost: $20,140.00
Annual Cost after first year: $2,600.00

Remark 1: vendor claims a 2-1 speed advantage over Oracle & Ingres
Remark 2: interfaces to spreadsheet, statistics, graphics & word processing
Remark 3: HLI a recent development for the VAX impremementation
Remark 4: price of graphics/spreadsheet from different vendor included in cost above.
Remark 5: !! REQUIRES database be off-loaded to add/delete attributes !!

EVALUATION CRITERIA CHECKLIST

MINIMUM REQUIREMENTS

MicroVax: Yes
VMS Operating System: Yes - only operating system supported
Relational Database: claims relational; also provides Hierarchical & Network Models
FORTRAN Interface: a recent addition

Query Language: Yes, employs a proprietary non-procedure query language

Product Support: Yes, consulting available
Notes on Data Types: All basic types supported
Memory | Disk Reqmts ~100MB disk: easily fits within constraints

OBJECTIVE FEATURES

Direct Support of Geographic Data Types: No direct support
Adequacy of File | Record Specs: Unlikely to be exceeded
Build In Menu Operation: Yes, menu generator available at extra cost
Communications Support: Yes - hardware dependent
Basic Statistical & Math Functions: Yes; plus a few additional scientific functions
Security System: minimal level of security - no password
On Line Help function: Yes
Approx. Five Year Cost: $30,540.00

SUBJECTIVE FEATURES

Group 1

Completeness of FORTRAN Interface: basic access to database records & tables
File manipulation ability: appears adequate
Max. Tables in Report Query...: Up to ten files in Report definition
Report Writer Avail; Flexiable: Yes, Unknown degree of sophisication
Store repeated queries: Yes, vendor claims all applications are compiled internally
Completeness of query language: appears comperable to SQL, maybe a little better
System recovery features: NONE

Graphical Interface: forms paint option available - menu generator
Business Graphics: interface (data transfer) to Tell-A-Graph
Spreadsheet function: interface (data transfer) to NCPCalc
Advanced statistical functions: interface (data transfer) to SPSS
Geographic based data functions: Not available

Speed of Storage & Retrieval: unknown - vendor claims superiority
Capacity for data: substantial
Ease of adding & deleting attributes: interrupts database use

Group 2

Import different data formats: read/write access to RMS files
Export different data formats: read/write access to RMS files
Completeness of on line help: unable to determine
Operating System interface: able to execute programs from within database system
Query log: no transaction logging
User log: Yes
Portability to other systems: poor
Cost of operational system:

ADDITIONAL INFORMATION AND SPECIFICATIONS

Other OS Supported: NONE
Query Procedures: PROPRIETARY
Transaction Logging: NO
Rollback | Recovery: NO
Forms Generator: YES
Report Writer: YES

ENHANCED ANALYSIS ENVIRONMENT

Integrated Spreadsheet: Interface to external
Built in Business Graphics: Interface ot external
Build in Advanced Statistics:
Interface to external GIS | GBF Interface: NONE identified

DATABASE SPECIFICATIONS

Max Char Per Field: 4,095 Max Char Per Record:

Max Records Per File: Unlimited Max Fields Per Record: 2,048

Max Files Per Database: Unlimited Max # of Indicies: Unlimited

Data Types: Byte, word, long, D_float, character, leading, trailing

IN FINALS: Y

Arizona Department of Transportation
PIMA Association of Governments Transportation Planning Division
Transportation Data Base Project

DataBase Name: Boeing RIM PHONE: (800) 551-0800 x151

Vendor: Boeing Computer Services
Representative: William H. Gillis

ADDRESS: P.O. Box 24346
City: Seattle State: WashingtZIP: 98124-0346

First Year Cost: $15,595.00
Annual Cost after first year: $1,875.00

Remark 1: 45 day on-site evaluation available, limied support for text data
Remark 2: Orientation is towards scientific & technical data
Remark 3: 60 commands available, stores data in machines format
Remark 4: Cost of graphics/spreadsheet from seperate vendor included in price above.
Remark 5: arrays & matricies available but limited to 1,021 words

EVALUATION CRITERIA CHECKLIST

MINIMUM REQUIREMENTS

MicroVax: Yes
VMS Operating System: Yes
Relational Database: Yes
FORTRAN Interface: Yes

Query Language: Yes, up to 10 arithmetic, logical conditions
Product Support: Yes, customization available
Notes on Data Types: Basic types plus arrays & matricies
Memory | Disk Reqmts ~100MB disk: Yes, 500 blocks of disk space for operation

OBJECTIVE FEATURES

Direct Support of Geographic Data Types: No direct support but array, matrix types available
Adequacy of File|Record Specs: completely adequate but not generous
Build In Menu Operation: NO! Also lacks an applications generator.
Communications Support: Yes, dependent on hardware
Basic Statistical & Math Functions: basic plus more scientific functions than other DBMS
Security System: Yes - password, but security only at table level
On Line Help function: Yes, can be expanded
Approx. Five Year Cost: $23,095.00

Evaluation Criteria Checklist (cont.)

SUBJECTIVE FEATURES

Group 1

of FORTRAN Interface: position in up to six tables supported
File manipulation ability: ref. up to 73 attributes in 1 statement
Max. Tables in Report Query...: 50 tables
Report Writer Avail; Flexiable: Yes, unknown level of sophisication
Store repeated queries: Yes - simple redirection of input to RIM
Completeness of query language: unknown - SQL like language
System recovery features: transaction logging & rollback/recovery

Graphical Interface: Not available
Business Graphics: Very limited
Spreadsheet function: No
Advanced statistical functions: No, but some advanced functions provided
Geographic based data functions: No but array & matrix access available

Speed of Storage & Retrieval:
Capacity for data: complete adquate
Ease of adding & deleting attributes: can be done with database on-line but is circuitous

Group 2

Import different data formats: any data which can be defined by std. fortran format
Export different data formats:---if type which can be specified by FORTRAN format statement
Completeness of on line help: unknown - unable to determine
Operating System interface: allows execution of programs within system
Query log: Not Available
User log: Not Available
Portability to other systems: available for CDC CYBER, VAX & IBM 43XX, 370, 308X, PRIME UNIVAC & HARRIS
Cost of operational system: $12,800

ADDITIONAL INFORMATION & SPECIFICATIONS

Query Procedures: SQL-like proprietry language

Transaction Logging: YES Rollback | Recovery: YES
Forms Generator: YES Report Writer: YES

ENHANCED ANALYSIS ENVIRONMENT

Integrated Spreadsheet: NO
Built in Business Graphics: YES
Build in Advanced Statistics: NO
GIS | GBF Interface: NO

DATABASE SPECIFICATIONS

Max Char Per Field: 1,021

Max Char Per Record:

Max Records Per File: 4 billion

Max Fields Per Record: 1,021

Max Files Per Database: Unlimited

Max # of Indicies: Unlimited

Data Types: integer, real, dbl precision, text, arrays, vectors & matricies

IN FINALS: Y

Arizona Department of Transportation
PIMA Association of Governments Transportation Planning Division
Transportation Data Base Project

DataBase Name: EMPRESS/32 PHONE: (416) 922-1743

Vendor: Rhodnius Incorporated
Representative: Diana Russof, Lee Robinson

ADDRESS:
City: State: ZIP:

First Year Cost: $14,050.00
Annual Cost after first year: $1,350.00

Remark1: user-defined functions supported - but for C only
Remark2: price of graphics/spreadsheet from separate vendor included in cost above.
Remark3: unstructured byte stream can be used for map image storage
Remark4:
Remark5:

EVALUATION CRITERIA CHECKLIST

MINIMUM REQUIREMENTS

MicroVax: Yes
VMS Operating System: Yes
Relational Database: Yes
FORTRAN Interface: Yes, also C & COBOL

Query Language: Yes - standard SQL

Product Support: Yes, customization available
Notes on Data Types: Yes, unstructed byte stream of special interest
Memory | Disk Reqmts ~100MB disk: Yes, ~3 MB disk, 1MB memory

OBJECTIVE FEATURES

Direct Support of Geographic Data Types: NO direct support, but has compatible structure
Adequacy of File | Record Specs: data limitations are not a concern
Build In Menu Operation: forms driven or command driven
Communications Support: DEC NET, VAX Cluster
Basic Statistical & Math Functions: AVG, SUM, MAX, MIN, COUNT; no variance or std. deviation
Security System: Yes - password plus SQL security provisions
On Line Help function: Yes
Approx. Five Year Cost: $19,450.00

Evaluation Criteria Checklist (cont.)

SUBJECTIVE FEATURES

Group 1

Completeness of FORTRAN Interface: can embed SQL type commands in programs
File manipulation ability: appears complete
Max. Tables in Report Query...: 9,999
Report Writer Avail; Flexiable: Yes, MWriter
Store repeated queries: Yes
Completeness of query language: standard SQL, complete
System recovery features: checkpoint / recovery utilities

Graphical Interface: Not Available
Business Graphics: Not Avaliable
Spreadsheet function: Not Available
Advanced statistical functions: Not Available
Geographic based data functions: Not Available

Speed of Storage & Retrieval:
Capacity for data: substancial - not a concern
Ease of adding & deleting attributes: appears no interference with db operation

Group 2

Import different data formats: supports bulk loading of ASCII files
Export different data formats: ASCII
Completeness of on line help: unknown
Operating System interface: Yes
Query log:
User log:
Portability to other systems: also available for UNIX
Cost of operational system: $10,800

Evaluation Criteria Checklist (cont.)

ADDITIONAL INFORMATION & SPECIFICATIONS

Other OS Supported: UNIX

Query Procedures: SQL

Transaction Logging: YES

Rollback | Recovery: YES

Forms Generator: YES

Report Writer: YES

ENHANCED ANALYSIS ENVIRONMENT

Integrated Spreadsheet: NO

Built in Business Graphics: NO

Build in Advanced Statistics: NO

GIS | GBF Interface: NO

DATABASE SPECIFICATIONS

Max Char Per Field: 2 billion

Max Char Per Record:

Max Records Per File: 2 billion

Max Fields Per Record: 9,999

Max Files Per Database: 9,999

Max # of Indicies: 1 billion

Data Types: long integer, float, variable length text, unstructed byte streams

IN FINALS: Y

Arizona Department of Transportation
PIMA Association of Governments Transportation Planning Division
Transportation Data Base Project

DataBase Name: InFoCen Relational DBMS PHONE: 703 442-8484

Vendor: 3CI
Representative: Scott Walker, Jr.

ADDRESS: 7500 Ambergate Place, Suite 3
City: McLean, State: VA. ZIP:

First Year Cost: $20,761.00
Annual Cost after first year: $2,066.00

Remark1: performance monitoring included
Remark2: better than others for handling textual data.
Remark3: price includes 3 days (1 person) training in Ft. Collins
Remark4: price above includes cost of spreadsheet from seperate vendor.
Remark5: database must be off-loaded to add / delete attributes !!

EVALUATION CRITERIA CHECKLIST

MINIMUM REQUIREMENTS

MicroVax: Yes
VMS Operating System: Yes
Relational Database: Yes
FORTRAN Interface: Yes, basic access provided

Query Language: Yes, Non-Structed Query Language

Product Support: Yes, customization available
Notes on Data Types: Basic types plus date, dollar & binary
Memory | Disk Reqmts ~100MB disk: ~4MB disk, 1M memory

OBJECTIVE FEATURES

Direct Support of Geographic Data Types: No direct support,
GIS Interface available for "MOSS" (USFS)
Adequacy of File | Record Specs: No limit
Build In Menu Operation: Yes
Communications Support: Yes - dependent on hardware
Basic Statistical & Math Functions: Yes
Security System: Yes, 6 security levels - password
On Line Help function: Yes
Approx. Five Year Cost: $29,025.00

Evaluation Criteria Checklist (cont.)

SUBJECTIVE FEATURES

Group 1

Completeness of FORTRAN Interface: basic access to files & records
File manipulation ability: appears complete
Max. Tables in Report Query...:
Report Writer Avail; Flexiable: Yes, limited flexibility
Store repeated queries: Yes
Completeness of query language: appears comperable to SQL
System recovery features: rollback / recovery
Graphical Interface: Not Available
Business Graphics: Yes
Spreadsheet function: Not Available
Advanced statistical functions: Not Available
Geographic based data functions: No, but has GIS Interface to MOSS
Capacity for data: essentially no limit
Ease of adding & deleting attributes: --- must shut down database

Group 2

Import different data formats: yes, "Fastload"
Export different data formats: Yes
Completeness of on line help: unable to determine
Operating System interface: YES
Query log: yes
User log: yes
Portability to other systems: Yes, VMS; DG MV AOS/VS
Cost of operational system:

ADDITIONAL INFORMATION & SPECIFICATIONS

Other OS Supported: VMS; DG MV AOS/VS, UNIX - BSD

Query Procedures: NSQL - proprietary

Transaction Logging: YES | Rollback | Recovery: YES

Forms Generator: YES | Report Writer: YES

ENHANCED ANALYSIS ENVIRONMENT

Integrated Spreadsheet: NO | Built in Business Graphics: YES

Build in Advanced Statistics: NO | GIS | GBF Interface: NO

Evaluation Criteria Checklist (cont.)

DATABASE SPECIFICATIONS

Max Char Per Field: 7,000 fixed 70,000 varMax Char Per Record:

Max Records Per File: Unlimited Max Fields Per Record: Unlimited

Max Files Per Database: Unlimited Max # of Indicies: Unlimited

Data Types: F_float, D_float,character, boolean, integer,queue?

IN FINALS: Y

Arizona Department of Transportation
PIMA Association of Governments Transportation Planning Division
Transportation Data Base Project

DataBase Name: INGRES PHONE: (416) 922-1743;
(800) 4-INGRES

Vendor: Relational Technology
Representative: Nicki Simpson

ADDRESS: 1080 Marina Village Parkway; P.O. Box 4006
City: Alameda State: CA. ZIP: 94501-9891

First Year Cost: $21,525.00
Annual Cost after first year: $3,225.00

Remark1: price includes training credits; spreadsheet from seperate vendor included in price above.
Remark2: Supportes multiple databases on a disk;
Remark3: supports SQL, QUEL & QBF query languages
Remark4: program interface allows SQL & 4GL commands within FORTRAN
Remark5: easy data transfer to PCs

EVALUATION CRITERIA CHECKLIST

MINIMUM REQUIREMENTS

MicroVax: YES
VMS Operating System: YES
Relational Database: YES
FORTRAN Interface: YES

Query Language: YES - SQL, QUEL & QBF

Product Support: YES
Notes on Data Types: Basic plus date & boolean
Memory | Disk Reqmts ~100MB disk:

OBJECTIVE FEATURES

Direct Support of Geographic Data Types: No
Adequacy of File | Record Specs: max. fields/record is a concern
Build In Menu Operation: Yes INGRES/MENU and applications generator
Communications Support: Yes - hardware dependent
Basic Statistical & Math Functions: Yes
Security System: Yes - imbedded in database
On Line Help function: Yes - can be expanded
Approx. Five Year Cost: $34,425.00

SUBJECTIVE FEATURES

Group 1

Completeness of FORTRAN Interface: substancial
Max. Tables in Report Query...: unlimited
Report Writer Avail; Flexiable: Yes, can be applied to views
Store repeated queries: Yes - option to optimize
Completeness of query language: standard SQL & QUEL; QBF is more limited but "user-friendly"
System recovery features: YES, rollback / recovery
Graphical Interface: Available with product for diffent vendors
Business Graphics: Yes, optional
Spreadsheet function: Not Available
Advanced statistical functions: Not Available
Geographic based data functions: Available from seperate vendor
Speed of Storage & Retrieval: an advantage
Capacity for data: max fields (127) per record is a concern
Ease of adding & deleting attributes: database is unaffected

Group 2

Import different data formats: Yes
Export different data formats: Yes, PCLink direcly supports export for PC software
Completeness of on line help: unknown - can be expanded
Operating System interface: Yes
Query log: Yes
User log: Yes
Portability to other systems: Yes
Cost of operational system: ~$14,000 + ~$2800/yr

ADDITIONAL INFORMATION & SPECIFICATIONS

Other OS Supported: VM/CMS, VAX/VMS & UNIX

Query Procedures: SQL, QUEL, Query by Forms

Transaction Logging: Yes Rollback | Recovery: Yes

Forms Generator: Yes Report Writer: Yes

ENHANCED ANALYSIS ENVIRONMENT

Integrated Spreadsheet: Not Available Built in Business Graphics: Yes

Build in Advanced Statistics: Not Available GIS | GBF Interface: Avail Seperately

Evaluation Criteria Checklist (cont.)

DATABASE SPECIFICATIONS

Max Char Per Field: 2,000

Max Char Per Record:

Max Records Per File: Unlimited

Max Fields Per Record: 127

Max Files Per Database: Unlimited

Max # of Indicies: Unlimited

Data Types: byte, word,long, D_float, F_float, char, boolean, date, G_float, H_float

IN FINALS: Y

Arizona Department of Transportation
PIMA Association of Governments Transportation Planning Division
Transportation Data Base Project

DataBase Name: Oracle PHONE: (415) 598-7538

Vendor: Oracle Corporation
Representative: Judy Anderson

ADDRESS: 20 Davis Drive
City: Belmont State: CA. ZIP: 94002

First Year Cost: $39,487.50
Annual Cost after first year: $4,387.50

Remark1: Becomes a resident process vs. an executable process
Remark2: user & Performance stats available; available for a variety of
Remark3: machines; compatible with IBM DB2
Remark4:
Remark5: price includes 5 days (1 person) training

EVALUATION CRITERIA CHECKLIST

MINIMUM REQUIREMENTS

MicroVax: Yes
VMS Operating System: Yes
Relational Database: Yes
FORTRAN Interface: Yes - precompile with embedded SQL provided

Query Language: standard SQL

Product Support: Yes - customization avail
Notes on Data Types: BASIC + Date, Boolean
Memory | Disk Reqmts ~100MB disk: approx 10MB disk min.2MB +
.5/user active memory

OBJECTIVE FEATURES

Direct Support of Geographic Data Types: No
Adequacy of File | Record Specs: adequate; more modest than Empress/32 or non-SQL systems
Build In Menu Operation: Yes - SQL MENU, Easy SQL, SQL Plus & SQL Forms
Communications Support: Yes - level of support hardware dependent
Basic Statistical & Math Functions: Yes but lacks standard deviation & variance
Security System: Yes - standard SQL security plus password
On Line Help function: Yes - extendable
Approx. Five Year Cost: $57,037.50

Evaluation Criteria Checklist (cont.)

SUBJECTIVE FEATURES

Group 1

Completeness of FORTRAN Interface: Precompile & embedded SQL
File manipulation ability: complete
Max. Tables in Report Query...: unlimited
Report Writer Avail; Flexiable: Yes, full support for views
Store repeated queries: Yes
Completeness of query language: standard SQL & more limited menu-driven Easy SQL
System recovery features: transaction logging, rollback/recovery
Graphical Interface: Not Avaliable
Business Graphics: Yes
Spreadsheet function: Yes - special version of 20/20 w/ access to tables
Advanced statistical functions: interface (data transfer) to SAS available
Geographic based data functions: Not Available
Speed of Storage & Retrieval:
Capacity for data: appears adequate
Ease of adding & deleting attributes: database operations unaffected

Group 2

Import different data formats: Yes
Completeness of on line help: unknown
Operating System interface: Yes
Query log: Yes
User log: Yes
Portability to other systems: Yes
Cost of operational system:

ADDITIONAL INFORMATION & SPECIFICATIONS

Other OS Supported: IBM VM/VMS CMS, UNIX, PC-DOS

Query Procedures: standard SQL

Transaction Logging: Yes Rollback | Recovery: Yes

Forms Generator: Yes Report Writer: Yes

ENHANCED ANALYSIS ENVIRONMENT

Integrated Spreadsheet: Yes Built in Business Graphics: Yes

Build in Advanced Statistics: Not Available GIS | GBF Interface: Not Available

DATABASE SPECIFICATIONS

Max Char Per Field: 240

Max Char Per Record:

Max Records Per File: Unlimited

Max Fields Per Record: 255

Max Files Per Database: Unlimited

Max # of Indicies: Unlimited

Data Types: byte, word, long,F_float,D_float, packed char, boolean, leading, trailing

IN FINALS: Y

APPENDIX B

EVALUATION QUESTIONS FOR THE
TRANSPORTATION DATABASE SYSTEM HOST SOFTWARE

File and Record Specifications

Please document the recommended minimum requirements for the program running on an multi-user (8 active) MicroVAX II under VMS.

Minimum real memory	- megabytes
add'l per user	- megabytes
Minimum Disk Space - for installation	- megabytes
for operation	- megabytes

Database Specifications:

Maximum Number of fields per record
Maximum Number of bytes per record
Maximum Number of records per table
Maximum Number of tables per database
Maximum Number of active databases

Can the database be spread over multiple disk drives, machines?

Can the database be backed up/restored from outside of the system or must be done from inside?

Can multiple databases reside on the same machine concurrently?

Data Manipulation

The following statements attempt to summarize essential characteristics in the manipulation of databases. Not all opera-

tions are covered and not all points are equally important. The responses to the following questions will be used to assess the data manipulation capabilities of the database.

1) Can subsets of tables be joined with other tables and subsets?

2) Are variable length fields available; more than one; can these fields be searched for keywords?

3) Can combined fields and/or transformation of fields be used as a key to a table?

4) Can indexes be added, deleted without affecting previously written programs/procedures?

5) Is a "semi-" distributed database version offered (i.e. a database which spans one than one machine) ?

6) Are null data values supported; Can a "not null" constraint be placed on a field or a query?

7) Are computed fields, dollar, long integer and date data types supported along with corresponding support functions?

8) Is data integrity supported with foreign keys?

9) Are procedural, as well as non-procedural commands, available in the interactive mode?

10) Are the results of a query available for browsing or must the query be repeated (longevity of result tables)?

Ease of Adding/Deleting Attributes

Which of the following best characterizes the database system.

1) Data must be "off-loaded" manually to add/delete attributes.

2) Data is automatically "off-loaded" but requires internal restructing of the database (database must be off-line or all processes suspended).

3) Attributes can be added/deleted while the database is live.

Security System and Reliability

Does the software provide integral data security including log-on control, file and/or field access control by user profile? Check all that apply.

1) Is security integrated into an **active** data dictionary?

2) What levels of access rights exist within the database system? (table, field, row - read, write, edit rights)

3) Are the security privileges available sufficient to complete hide the existence of a field from those not authorized?

4) Does the database provide transaction logging?

5) Automatic rollback/recovery?

6) a two-phase commit for updates?

"MENUING" Capability

Does the software provide build-in tools to create (using a build-in language or a FORTRAN interface) menu driven applications?

1) Are menus (or verbose prompts) available for some modules?

2) Are menus or prompts available for essentially all major operations (edit, report, select,...) and user log-on automatic menu option available?

3) A menu generator provided to allow development of menu driven procedures without use of the Host Language Interface? IF yes, does it contain or allow procedural constructs.

Communications Support

Does the software provide support for establishing and maintaining remote communications and provide full access by remote user with the appropriate equipment? Check all that apply.

1) What level of transmission is supported (i.e. baud rate)?

2) Is remote access **controllable** from within a program?

3) Is an assembly language interface available?

Store Repeated Queries/Procedures

Respond as appropriate.

1) Can time consuming operations be set up to run unattended?

2) Is a history or listing operation provided to allow correction or later re-issue of queries or commands?

3) Is a means provided of storing queries/procedures which are to be repeated (stored macro facility or command files)?

4) Is any gain in efficient available for repeated queries without special programming or building of indexes (i.e. re-use of query strategy)?

5) Can stored queries/procedures call other stored procedures?

6) Is a means provided to store a value for later reference?

7) Are IF..THEN, CASE.. WHILE..DO constructs available for use in procedures/programs?

Completeness of the programming language interface for FORTRAN

Respond as appropriate.

1) Only the basic operations are available - Open/Close File, Open/Close Index, Get/Put Record,

First/Last/Next/Prev Record.

2) Additional commands are provided but use of the non-procedural data manipulation language/data definition language is **not provided.**

3) Embedded SQL-type commands are supported in the FORTRAN interface.

4) A pre-compiler as well as "run-time" database system calls are provided.

Support Geographic Data Types

Does the software package provide build-in support for geographic data types (i.e., point, line, polygon, area)? Check one of the following assessments.

1) No direct support of geographic data type or structure suitable for storage of map images.

2) No direct support, but array, bit image or other data structure(s) available for storage of map images or information.

3) Programmed interface to external Geographic Information System (GIS) or Geographic Base File system (GBF).

4) Full blow GIS or GBF is available within DBMS.

Geographic Based Data Functions

Respond as appropriate.

1) Does the DBMS provide an aggregation command to create a table summarizing another table?

CREATE ____ (field1, @SUM (field2), AVG(field3))
FROM _____ GROUP BY field1

2) Is a command available to (automatically) look-up the value in a field and substitute a value based on the value in the "control" table? Is a table look-up function provided?

3) Is a table disaggregation function provided (the reverse of the aggregation function on item 1 above)?

BASIC statistical and mathematical functions

Does the software package provide all of the following mathematical and statistical functions: +, -, /, *, MIN, MAX, AVG, COUNT, SUM, Standard Deviation, Variance? Check all that apply.

1) All of the above operators provided.

2) Additional mathematical and/or financial operators provided (i.e. Ln Log10, EXP, SQRT, PMT, IRR,...).

3) special functions are provided to support strings & date data types.

Completeness of the QUERY Language

Check all that apply.

1) Allows computed values and multiple criteria in select

statements.

2) Allows queries/subsetting of virtual tables.

3) Allows multiple query/subset (virtual tables) to exists for a user.

Formatted Output / Report Writer

Check all that apply.

1) Is a report writer available for developing output formats?

2) Can report formats be saved, shared and re-applied?

3) Does the report writer support at least 5 levels of sorting with break points, detail suppression, headers at break points?

4) Can the reports be applied to virtual tables?

5) Is the report writer screen oriented or menu driven rather than command oriented?

6) Mathematical and/or Summary Operators available within the report writer at break points?

7) Can the database support (either thru the report writer or other non-programmed facilities) user-defined data export formats?

Graphical Interface

1) Does the GIS(GBF)/DBMS combination support a graphical mapping user interface to identify areas of interest?

Business Graphics

1) Are standard business graphics available within the DBMS (pie, bar, line, scatter)?

2) Are additional graphics options provided (font control, hi-lo charts, 3-D graphics,...)?

3) Can the graph specification be saved and can it be applied to virtual tables?

Spreadsheet

1) Is a full feature spreadsheet available within the database system?

2) Are tables of the database accessible from the spreadsheet?

3) Is the size of the spreadsheet sufficient to allow a substantial size table (5000+ records) to be loaded and does it check for capacity prior to loading or processing?

Advanced Statistical Functions

Check all statistical groups/functions which are provided

1) Descriptive Statistics & Frequencies

2) Cross-tabulation & Correlation Procedures

3) Single and Multiple Variable regression (linear & non-linear)

4) Aggregate Analysis & Univariate

5) Are time-series, ANOVA, MANOVA procedures provided

6) Discriminant Analysis

Speed of Storage / Retrieval / Operation

1) Are functions provided for the system administrator to evaluate and optimize the DBMS operation?

2) Is a "fast load", "bulk load" facility provided?

Operating System Interface

Check all that apply.

1) Can invoke executable programs from within system?

2) Can a "seamless" mixture of DBMS and non-DBMS procedures be provided from a menu oriented system (i.e., user selects option, not obvious which type procedure is run, user only sees the results of the operation)?

On-Line Help Function

Does the software provide On-line help within the standard interface? Are facilities for programmed help within a programmed menu application? Check all that apply.

1) Is some on-line HELP provided beyond prompts?

2) Is on-line help always available (all modules at all times)?

3) Is the HELP function modifiable and/or extendable and linkable to user-developed menu modules?

Direct Support of Different Data Import and Export Types

Check all that apply.

1) ASCII

2) ASCII and DIF or other popular data format supported

3) ASCII, DIF (spreadsheet) and direct support of system binary files (RMS direct, sequential,...)

4) Allow editmask or select in the importation of data

5) Allows export of data with format control within system without programming

6) Allow column selection in importing

Query Log

Check if the statement is true of the dbms.

1) Maintains a history of procedures, "programs" used and the number of runs and time, i/o requirements, query successful (I/O per program, type of access requests, CPU utilization, inadequate resource availability, overhead of the monitor.

User Log

Check if the statement is true of the DBMS.

1) Maintains a history of user log-ons, table uses and activities.

2) Does the system maintain information on database activities conducive to fine-tuning the performance?

License, Purchase Agreement COST

What is the cost of the base system and specified options for an 8+ user microVAX II installation:

	INITIAL PRICE	ANNUAL FEES
BASE DBMS Cost		
Fortran Interface		
Forms Writer		
Report Writer		
Menu System (if available)		
Spreadsheet (if available)		
Statistics (if available)		
Business Graphics (if available)		
Mapping, GIS or GBF (if available)		

Portability

Can the file structures (schema), procedures (both programmed & saved) be readily transferred to other systems?

1) Is the DBMS available for other popular operating systems, e.g. UNIX, VM, PC-DOS?

2) Are the constructs and terminology used sufficiently standard that the database design and applications can be readily transferred to a different database system?

Any additional information, representative benchmarks or comparisons with competing packages which highlights beneficial, distinguishing characteristics of the database system would be most appreciated and welcome.

APPENDIX C

INGRES

INGRES is a business oriented relational database which features three different query mechanisms, fast operation and a user-friendly interface. It is the principal product of Relational Technology, Inc. and has been available commercially since 1981. Relational Technology maintains headquarters in Alameda, California and has fifteen division offices nationwide and supports INGRES for three major operating environments, DEC VAX/VMS, IBM VM/CMS and UNIX.

INGRES is in now in its fifth major version and documents a record of consistent performance improvements with each major release. According to the company, INGRES consistently wins industry benchmarks for its speed. Relational Technology has a reputation for providing excellent support for INGRES and Datapro rated INGRES as among the best supported products. In a recent review of INGRES and its distributed database component, INGRES/STAR, Digital Review found the support to be laudable, but also considered the documentation for system administrators a detraction from INGRES' "near-stellar" performance.

INGRES is best characterized as a "pure" relational database. This contrasts with a number of database products which have added the relational banner to their network or hierarchical based products. As such, INGRES provides complete tools for data management within the relational framework. This includes the management of security through the data dictionary, the modification of the database schema using the its relational operators and full support of null values and views (virtual tables in the database system).

The standard data types are supported including integer, real, double precision real numbers, character and variable length text. In addition, date and boolean data types are provided. The text handling facilities are somewhat better than that of Oracle but are modest compared to that of InfoCen. The comparatively limited number of fields (127 per table) is an area of concern but is somewhat offset by an unlimited number of tables per database.

Relational Technology has continually enhanced INGRES and is among the leaders in the development of a true distributed database (a fully accessible and modifiable database spread over multiple machines and operating systems). While this may not be immediately useful to PAGTPD, it provides for future implementation of a shared database with other local governmental authorities and the Arizona Department of Transportation. The availability of a standard query system also contributes to this potential.

Query facilities are among INGRES' strong points. It supports three different query interfaces, Structured Query Language (SQL - pronounced SEQUEL), QUAL and Query-By-Forms (QBF), each of which is optimized prior to execution. SQL and QUEL are very

similar and complete query languages and are basically alternatives to one another. SQL is more popular, considered more "user-friendly" and is quickly becoming the standard language for database data manipulation. QUEL is considered by some to be more complete but is also more terse in its expression. The two languages may be used interchangeably.

The Query-By-Forms interface provides menu-driven access to the database without having to learn either SQL or QUEL. This would likely be the first choice of occasional user and provides access comparable to that available with the complete query facilities of some other databases. There is a performance penalty involved in using the QBF interface for complex operations. Queries which are to be repeated may be saved in a "script" file and INGRES provides an option to gain some efficient by reusing the same strategy to execute the query.

Support for SQL allows applications to be portable across different hardware and software environments and enhances the portability of the transportation database design and procedures. As with Oracle, INGRES' implementation of SQL is compatible with the ANSI SQL standards. As a "true" relational system, INGRES allows queries on views unlike Boeing RIM and some other systems investigated.

INGRES' inherent analytical capabilities are relatively modest. The built-in functions include count, min, max, average, sum, standard deviation and variance as do many DBMS. As with some other database systems, user defined functions can be added but must be programmed in C. There is no build-in statistics or spreadsheet function. INGRES does provide a business graphics capability which is accessable within its menu-driven interface. Standard business graphics (bar graph, line, x-y and pie charts) are supported. Like the other databases considered in this review, INGRES does not have a built-in GIS or GBF capabilities. However, two major vendors of GIS systems contacted during the course of this investigation provides well-integrated database access using INGRES.

INGRES provides a well developed interface to programming languages on the VAX specifically including FORTRAN. In addition to the more common calls to database functions, INGRES supports embedded SQL and 4GL commands into any of seven programming languages. Data exchange with external applications is accomplished by producing a standard system file containing the data of interest. This can be accomplished using the either of the report writer options or can be programmed using the FORTRAN interface. A user-friendly module (PCLINK) is available for access to the database from personal computers and can produce files suitable for 1-2-3, dBASE and other application programs.

INGRES is also distinguished by its ease of use providing menu-driven access to all database operations and on-line help. The menu-driven modules provided include: Query-By-Forms, Applications-By-Forms, Report-By-Forms, VIFRED and VIGRAPH. The

Applications-By-Forms allows development of procedures (including menu-driven applications) without conventional programming. The Report-By-Forms is a menu-driven report writer which serves as an alternative to the more complete Report Writer module. VIFRED is a screen/menu oriented forms design module used to develop customized input and display screens. The full range of display characteristics and validation checks can be specified using VIFRED. VIGRAPH is the menu oriented graphics module which is used to develop customized graphics. In all of these processes as well as in interactive QUEL and SQL, on-line help is available. The HELP function uses an INGRES database and can be expanded to include help for user developed applications.

Oracle

Oracle is a popular business oriented relational database which features a near complete implementation of the relational model, widespread availability and compatibility with industry standards. Oracle was first introduced in 1979 and is distributed by the Oracle Corporation. Oracle Corporation maintains 14 regional offices in the United States and has its headquarters in Belmond, California. Oracle is in now in its fifth major version.

Oracle is best characterized as a "pure" relational database in the same view as INGRES. Complete tools for data management within the relational framework. This includes the management of security through the data dictionary, the modification of the database schema using the its relational operators and full support of null values and views (virtual tables in the database system). Oracle also provides password protection which is not available with Accent R, Boeing RIM or INGRES. Oracle has an unsubstantiated reputation as a solid if unspectacular performer in data manipulation.

The standard data types are supported including integer, real, double precision real numbers, character and variable length text. In addition, a boolean data type and a variable length text data type are provided. While variable length text fields are supported, the relative small field size (maximum 240 characters) limits its usefulness. The number of fields per record (255) is more generous than that available with INGRES but is much all than is available in some other systems. Another area of concern is the maximum number of tables in a database. Digital Review cited a maximum of 64 tables per database which would be very limiting for a relational database system. When questioned, Oracle Corporation indicated that the number of tables is unlimited.

Oracle Corporation is among the leaders in the development of a true distributed database (a fully accessible and modifiable database spread over multiple machines and operating systems) and currently supports retrieval access to DB2 databases on IBM equipment.

Support for the Structured Query Language (SQL) has long been a strong point for Oracle since the company introduced the first implementation of SQL in 1979. Oracle Corporation also provides a user-friendly, menu-driven environment, EASY*SQL for occasional users. As with the INGRES user-friendly shell for SQL, there appears to be a performance penalty for complex queries and more limited access to the database. This would likely be the first choice of occasional user and provides access comparable to that available with the complete query facilities of some other databases. Oracle's support for SQL allows applications to be portable across different hardware and software environments and enhances the portability of the transportation database design and procedures. Support for views is included in both query

environments.

Oracle' inherent analytical capabilities are stronger than that of INGRES primarily because a spreadsheet module is build-in and provides access to tables in the database modest. The built-in functions include count, min, max, average, sum but excludes standard deviation and variance. There is no build-in statistics. Oracle does provide a business graphics capability which is accessable within its menu-driven interface. Standard business graphics (bar graph, line, x-y and pie charts) are supported. Like the other databases considered in this review, Oracle does not have a built-in GIS or GBF capabilities. Such a capability could be developed with the Host Language Interface to a FORTRAN program.

Oracle provides a well developed interface to programming languages on the VAX specifically including FORTRAN. In addition to the more common calls to database functions, Oracle supports embedded SQL commands into any of several programming languages. This ability is comparable to that of INGRES and Empress/32 and is more substantial than that of the remaining database systems.

Oracle is also distinguished by its ease of use providing menu-driven access to all database operations and on-line help. SQL MENU is the primary menu system with additional menu-driven modules available: EASY SQL, SQL GRAPH, SQL CALC, SQL FORMS and SQL PLUS. EASY SQL is the menu driven shell for access to the database without using SQL. SQL GRAPH is the menu oriented graphics module which is used to develop customized graphics. SQL Calc is the previously mentioned spreadsheet module which provides access to tables within the database from a 1-2-3 like spreadsheet. SQL FORMS is the forms generator and SQL PLUS is the report writer module. Both modules are believed to be comparable to the corresponding functions in INGRES. In all, Oracle provides an environment which is substantially more "user-friendly" than the other systems considered excepting INGRES. The HELP function uses an Oracle database and can be expanded to include help for user developed applications.

ACCENT R

ACCENT R is a fourth generation programming language and data base management system that uses a non-standard query language. It is distributed by National Information Systems, Inc. (NIS) of Cupertino, CA. NIS claims that ACCENT R is twice as fast as Oracle and INGRES in input/output and CPU time. They also claim that ACCENT R applications are far more easier to modify than INGRES applications.

File structure specifications and limitations for ACCENT R include a maximum number of fields per record of 4026 and a maximum number of characters per field of 30,000. There is no limit on the number of records per file. The system requires up to 100MB of storage and files can be stored in either ASCII or binary format. ACCENT R supports all of the standard data types along with computed fields and date type. It does not, however, support a null value for a data field. A major drawback to ACCENT R is that the data base must be taken "off-line" or "off-loaded" before any additions or deletions can be made to the data base structure. Also, ACCENT R employs standard VMS file security, it does not have its own built-in security system. This means that data can only be "secured" at the file level. Additionally, ACCENT R does not provide any recovery or roll-back procedures if the current version of the data base is damaged or destroyed.

ACCENT R employs a non-standard query language that they claim is more complete than SQL. It allows computed values and multiple criteria in select statements. Tables created through queries cannot be held in memory, they must be written to disk to be used later. This may cause problems with disk space if temporary files created by these queries are not erased on a periodic basis. Repeated queries and operations can be stored in macro or batch files. These stored queries can call other stored queries. Advanced programming constructs such as IF..THEN..ELSE and WHILE..DO are available.

ACCENT R is command driven, but creating menus is supposedly easy with DataPaint utility. A "seamless" mix of ACCENT R and FORTRAN (or other non-ACCENT R) applications can be developed. ACCENT R's Host Language Interface does support FORTRAN, but it does not allow full use of query language within FORTRAN. A call can be made to ACCENT R from FORTRAN and parameters passed between them.

ACCENT R's report writer is user-friendly and allows full screen design of forms that can be saved for later use, but it is not as flexible as the report writers of INGRES, Oracle, and Empress/32. Report writer also allows for page subtotals, subtotals at break points, headings, and footnotes. ACCENT R does not have a spreadsheet built-in, but can export data to an outside spreadsheet package. ACCENT R does not have a built-in graphics package or graphical interface, but NIS states that at least one private consulting firm has written code to allow ACCENT R to interface with a stand-alone graphics package.

The base price for ACCENT R is $11,500 plus $1,000 for an end user interface and $2,400 for the Host Language Interface. There is also a one-time start up cost of $950 that includes documentation and hotlines with quick response technical support. The initial license fee is $1490 with an annual fee of 15% of the license fee.

InFoCen

InFoCen is a relational database management system and fourth generation language that employs a proprietary Non-Standard Query Language (NSQL). It is distributed by 3CI of Fort Collins, CO. 3CI has a representative in the Washington area, it is not known if their is a representative in the Tucson area.

All of the basic file structure specification maximums are unlimited. These include characters per field, fields per record, records per table, tables per data base, and the number of active databases. Memory requirements for InFoCen are 3 MB of real memory and 2 MB of disk space for an installed system. Databases can be stored over multiple volumes, and multiple data base systems can reside and be used at the same time on the same volume. InFoCen supports all of the standard data types plus dollar and date types. It also supports a null data value. The main negative feature about InFoCen is that databases must be "off-loaded" for the data base structure to be modified. InFoCen does have a fast-load feature which quickly reloads the data base. Variable length fields are also supported and can be used in text searches. Security is integrated into the active data dictionary and up to 44 items, such as fields and records, can be used as security criteria.

The query language used by InFoCen is non-standard as opposed to ANSI approved SQL. It allows the use of computed values and multiple criteria in query selects. Virtual tables cannot be used for further queries unless they have been saved to a disk file. This may cause a disk space problem if multiple queries are performed consistently. The query language commands can be saved into macros if they are to be run repeatedly. Macros can be used to program mathematical functions that are not supported by the DBMS. Macros are also recursive (i.e., a macro can call itself).

InFoCen is a command driven language that does not have a full screen menu generator. Menus can be created using macros, but that involves in-depth programming instead of "screen painting". The DBMS supports an on-line help system that can be modified to add user created procedures and further explanations to current help messages. The FORTRAN interface allows the embedding of query language commands in FORTRAN code.

InFoCen does not support geographic data types, but data can imported/exported using an outside GIS. The report writer can save and reuse reports. It also subtotals by page and break points. Files can exported from InFoCen to Lotus, R:base 5000, and DIF structures. Although the DBMS does not contain a spreadsheet, data can be passed to an outside one easily. A version of InFoCen for 80386 based IBM-PCs will be available in the fall of 1987.

The base price for InFoCen is $15,895 with an annual 13% users fee. This includes one complete set of documentation and 3

updates per year. For an extra cost, two people will be trained at InFoCen's Fort Collins, CO headquarters.

Empress/32

Empress/32 is a relational database management system that works along with a fourth generation applications builder, M-Builder. It employs SQL as its query language. Empress/32 is distributed by Rhodnius Incorporated of Toronto, Canada.

All of the basic file structure specification maximums are unlimited. These include characters per field, fields per record, records per table, and tables per data base. Memory requirements for the system are 640 KB RAM, 5 MB of disk space for installation and 3MB of disk space for operation. At the current time, data bases cannot stored on multiple volumes, which does place a physical limit on the maximum size of a data base. Empress/32 supports all of the standard data types along with dollar, date, time, long integer, and bulk data types. The bulk data type can be very useful for storing map/geographical characteristics. Variable length fields are also available, either as text or bulk data. A positive feature of Empress/32 is that the data base structure can be modified while the data base is on-line, or "live". The 4GL applications package, M-Builder, does have procedural as well as non-procedural commands.

Empress/32 is command driven but menus can be created through the use of M-Builder. "Screen painting" is employed to created menus as well as design forms, which is very helpful. This is done using two utilities, M-Screen and M-Writer. Another good feature of Empress/32 is that remote access devices, such as a serial port, can be controlled from within M-Builder. The host language interface does support FORTRAN, Cobol, and "C". SQL commands may be embedded in FORTRAN. New commands can be developed and implemented by the user, but only through the use of a user defined "C" extent.

The query language used by Empress/32 is SQL. Like the other packages, SQL code can be saved and run in a batch mode. Nested selects can also be performed. Empress/32 does not have any advanced statistical functions, although it does support MIN, MAX, AVG, but not standard deviation and variance. Empress/32 also does not have business graphics or spreadsheet capabilities. On-line help is available for all of the component of Empress/32, including SQL, M-Builder, M-Screen, and M-Writer. The help function can be modified by the user.

The base price for Empress/32 is $9,000 which includes one set of documentation and a three month warranty. The annual maintenance fee is $1,350. Rhodnius will provide a week of training for an additional $1,500.

Boeing RIM

Boeing RIM is a relational data storage and retrieval system, designed mainly to support engineering and scientific data handling requirements. Boeing RIM employs an SQL-like language as its data base query language. Boeing RIM is distributed by Boeing Computer Services of Seattle, WA.

Boeing RIM is designed mainly as a scientific file handler, and it may not be flexible enough to accomodate all of PAGTPD's needs. There are physical limits placed on the number of fields per record, the bytes per record, and the number of data bases one user can have active at one time. The standard data types are supported by Boeing RIM, along with date and time. Attributes can be added/deleted while the data base is on-line, which is a plus.

The query language used by Boeing RIM is SQL-like, but not entirely SQL compatible. Frequently performed queries can be set up to run in batch mode. Stored queries may also call other stored queries, which eliminates duplication.

One drawback to Boeing RIM is that it does not have a security or file access system which prohibits users from accessing specified information in a file or record. Another problem with this DBMS is that it is command driven, with no 4GL available to create menus and screens. Basic file manipulation commands can be used in FORTRAN through the host language interface, but these commands are limited. The report writer is capable of meeting all of the basic needs of generating reports and special file outputs, but it command driven, which makes designing special forms and screens very programmer intensive. It is not known whether report formats can be applied to virtual tables created through queries.

Even though Boeing RIM is command driven, full screen data input and editing is supported. The DBMS also supports very limited business graphics. No built-in spreadsheet or advanced statistical functions are supported. The DBMS does provide on-line help for all of its components. The help function can be modified by the user.

APPENDIX D

ALGORITHM
A set of well-defined rules for the solution of a problem in a finite number of steps.

APPLICATION ENVIRONMENT
The application or group of applications for which DBMS implementation is under consideration.

APPLICATION GENERATOR
A powerful database application development tool designed to simplify and reduce on-line application development time. These tools generally provide for: screens forms generation; database creation; conversational program generation; testing/debugging; system documentation; and computer-aided instruction.

APPLICATION PROGRAM
A program written for or by a user that applies to a particular application.

BASIC ARITHMETIC OPERATOR
Any of the following four symbols, used for computation:
"*" multiplication
"+" addition
"-" subtraction, or
"/" division

BATCH
Describes a mode of processing where computer programs execute from start to finish without communicating with the user; contrast with INTERACTIVE.

BATCH PROCESSING
The processing of data or the accomplishment of jobs accumulated in advance in such a manner that each accumulation thus formed is processed in the same run.

BINARY FILE
A file that the computer can process consisting of only zeros and ones.

BOOLEAN
A data item, value or expression which yields a logical true or false.

BOOLEAN OPERATION
Any operation in which each of the operands and the result take one of two values. An operation that follows the rules of boolean algebra.

BUILT-IN FUNCTION*
Any of the computational or statistical operations available in a computer program without programming.

CALL
The action of bringing a computer program or subroutine into effect usually by specifying the entry conditions and jumping to an entry point.

CHARACTER DATA
Data in the form of strings of letters, digits, punctuation marks, and spaces that cannot be operated on mathematically.

CHARACTER FIELD
A field whose values will be treated as character data.

DATABASE
A collection of tables which may share fields among tables and are interrelated.

DATABASE ADMINISTRATOR
A person or persons given the responsibility for the definition, organization, protection, and efficiency of the database for an enterprise.

DATABASE INDEX
A file containing information specifying the sequence for accessing records within a table without regard to the physical ordering of records in the table.

DATABASE SCHEMA
Information which defines the data items within a database, the logical or physical grouping of items, the relationships among groups and the key items used to organize the data. The schema may also define the validity checks which are applied to the data.

DATABASE MANAGEMENT SYSTEM
A system that allows multiple independent users to have concurrent access to a central repository of information.

DATA COMPRESSION
A technique that saves storage space by eliminating gaps, empty fields, redundancies, or unnecessary data to shorten to length of records or blocks.

DATA DEFINITION
A program statement that describes the features of, specifies relationships of, or establishes context of data.

DATA DICTIONARY
A file containing a listing and description of the tables and fields which comprise a database.

DATA FILE
A collection of related data records organized in a specific manner.

DATA INTEGRITY
The concept that all units of data must be protected against accidental or deliberate invalidation.

DATA MANAGEMENT
The function of controlling the acquisition, analysis, storage, retrieval and distribution of data.

DEBUG
To detect, trace, and eliminate mistakes in computer programs or in other software. Synonymous with checkout.

DIRECT ACCESS
A means of locating database records directly without examining every record in the table

END-USER
The ultimate source or destination of information flowing through a system. An end-user may be an application program, an operation, or a data medium.

FIELD
One piece of information within a record which is defined for each record in a table; the information is defined by its name, format, length, type, etc. and may be qualified by validity and integrity checks; also known as an "attribute."

FORMAT
The arrangement or layout of data on a data medium.

FUNCTION
A mathematical entity whose value depends in a specific manner on the values of one or more independent variables, not more than one value of the dependent variable corresponding to each permissible combination of values from the respective ranges of the independent variables.

GEOGRAPHIC BASE FILE SYSTEM
A user-oriented system designed to create, edit and maintain a standardized geographic reference capability in a database management system environment. This system provides a framework for geographic inquiry and on-line geocoding.

GRAPHICAL INTERFACE
A user-interface which emphasizes the geographic aspects of the information presented. Such an interface would typically be used in a geographic information system or in interactive mapping software where the user is allowed to specify an area of interest by selecting one or more parts of a map displayed on a graphics terminal.

HOST LANGUAGE INTERFACE
A programmed interface between a database system (or other program) and a standard programming language such as FORTRAN. Typically, a HLI interface provides a means of accessing the data from within the database without reforming the data. At a minimum, the interface should provide facilities for accessing and modifying any data item of a specified record, deleting an existing record or creating a new data record.

INDEX
A table which maps each data item of a specific column into the remaining data of the same row. A table which maps an identifier of an entity into one or more attribute values of that entity.

INPUT DATA
Data being received or to be received into a device or into a computer program.

INPUT VALIDATION
The process of ensuring that the values entered into a database field are consistent with its field characteristics.

INQUIRY
A request for information from storage.

INTERACTIVE
Describes a mode of data processing where the user is prompted to pass information to a program as it executes.

INTERFACE
A shared boundary. An interface might be a hardware component or a portion of storage accessed by two or more computer programs.

JOIN
A database operation used to create a table from two or more tables having a command element. Information from either of the two original tables may then be accessed for reports, queries or other operations.

JOURNAL
A record of all environmental conditions and changes relative to the database. It may include time and date stamps, user identification, attempted security breaches, changes to a database, etc.

JOURNALING
Recording transactions against a database so that it can be reconstructed by applying the transactions in the journal against a previous version.

KEY
An entry in the database index derived from one or more fields in the database.

LOGICAL RECORD
A collection of one or more related data values.

LOGICAL RELATIONSHIP
The relationship that exists between two units of logical data.

MACRO
Stored text that has been given a name and can be used a input to a computer program as a substitution for a command, value or another macro.

MAIN STORAGE
Program-addressable storage from which instructions and other data can be loaded directly into registers for subsequent execution or processing.

MANAGEMENT INFORMATION SYSTEM (MIS)
Management performed with the aid of automatic data processing. An information system designed to aid in the performance of management functions.

MENU-DRIVEN
A process that allows the user to select a course of action based on a limit set of options presented; the implication is the options available are always limited to a managable number and are presented to the user in a pre-defined sequence.

NULL VALUE
A value which can be applied to all field types, can be distinguished from missing values and is used to denote the lack of information for a field.

NUMERIC FIELD
A field containing a number that can be used in computations.

ON-LINE
Pertaining to the operation of a functional unit that is under the continual control of a computer. Also used to describe a user's access to a computer via a terminal.

OPERATING SYSTEM
Software that controls the execution of computer programs and that may provide scheduling, debugging, input/output control, accounting, compilation, storage assignment, data management, and related services.

OUTPUT DATA
Data being delivered, or to be delivered from a device or from a computer program.

PASSWORD
A unique string of characters that a program, computer operator, or user must supply to meet security requirements before gaining access to data.

PHYSICAL RECORD
A record whose characteristics depend on the manner or form in which it is stored, retrieved, or moved. A physical record may consist of all or part of a logical record.

PROMPT
In an interactive environment, any symbol or message to which a user is expected to respond.

QUERY
A search of the database for only those records that meet specified conditions.

QUERY LANGUAGE
A language for the terminal user to retrieve and update data in the managed database.

RECORD
The basic unit for organizing data in a table; a record is divided into fields and stores information for each subject in the database (for example, each record might hold information about one employee).

RECOVERY ROUTINE
A routine that is entered when a error occurs during the performance of an associated operation.

RELATIONAL DATABASE SYSTEM
A database management system whereby the database consists of one or more rectangular tables of records and fields in different files that are linked (cross-referenced) by the content of fields that are common to the different files.

REPEATING FIELD
A field that can contain more than one field value, each of which is called an element.

REPORT
The output of a query either displayed at the terminal or printed on paper.

REPORT GENERATOR
A user-oriented system designed to create hard copy reports from the database. A report generator usually provides ease of use, flexible output formats, editing, sorting, sub-totals and a variety of other features.

SEARCH
The examination of a set for one or more items having a given property.

SECURITY
The limitation of user access to all or part of the data in a database, and to some commands.

SELECTION CRITERIA
Specifications that identify a subset of records from a table in the database to be the subject of a query or other processing.

SEQUENTIAL ACCESS
An access method by which the database is searched record-by-record until the desired record is found.

STATISTICAL ANALYSIS
A user-oriented system containing a variety of statistical programs for the analysis of data. Programs generally include routines for descriptive statistics, correlation, multiple regression, discriminate analysis, scatter diagrams, factor analysis, and many more.

STATISTICAL MAPPING
A variety of low-cost programs designed to prepare computer maps of statistical data on line printers and computer plotters. Most mapping software provide for isopleth (contour), choropleth (district) and incident mapping.

TABLE
An array of data each item of which may be unambiguously identified by means of one or more arguments.

TEMPORARY TABLE
A table that exists only for the duration of the session in which it was created.

TRANSACTION
In on-line systems, an exchange between a terminal and another device that accomplishes a particular action or result.

VARIABLE LENGTH FIELD
A field with a variable number of position to store characters, using only as many positions as it needs.

VIEW/VIRTUAL TABLE

A specific collection of data items from the database which is accessible by the user and may be treated as a table but which does not physically exist in the manner presented. A view may often be pre-defined in the data dictionary or may be created as a result of a user command. The specified collection of data items may be queried, manipulated or operated on as a standard table in the database and may be converted to a physical table by a user command.

VIRTUAL DATA

A logical unit of data that is materialized and does not exist as a physical unit of data.